

GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: January 28, 2004, 09:19:15 ; Search time 46 Seconds
(without alignments)
1880.565 Million cell updates/sec

Title: US-09-978-242-1

Perfect score: 2898

Sequence: 1 MATEHVNNGTSPMDTTSA.....RGPGSGARGVRAGKRGGRS 545

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 1107863 seqs, 158726573 residues

Total number of hits satisfying chosen parameters: 1107863

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

A_Geneseq_19Jun03.*
1: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1980.DAT.*
2: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1981.DAT.*
3: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1982.DAT.*
4: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1983.DAT.*
5: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1984.DAT.*
6: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1985.DAT.*
7: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1986.DAT.*
8: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1987.DAT.*
9: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1988.DAT.*
10: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1989.DAT.*
11: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1990.DAT.*
12: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1991.DAT.*
13: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1992.DAT.*
14: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1993.DAT.*
15: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1994.DAT.*
16: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1995.DAT.*
17: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1996.DAT.*
18: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1997.DAT.*
19: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1998.DAT.*
20: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1999.DAT.*
21: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA2000.DAT.*
22: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA2001.DAT.*
23: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA2002.DAT.*
24: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA2003.DAT.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length DB	ID	Description
1	2898	100.0	545	20 AAY24418	Human nucleolin-11
2	2898	100.0	545	23 ABG32846	Human nucleolin-11
3	2864	98.8	579	21 AAB56837	Human prostate can
4	2863	98.8	623	21 AAY70242	Human RNA-associat
5	2863	98.8	704	22 AAG74545	Human colon cancer
6	2860	98.7	561	23 ABB77791	Amino acid sequenc
7	2806	96.8	762	22 ABG04610	Novel human diagno
8	2355	81.3	633	21 AAY94961	Human secreted pro
9	1647.5	56.8	397	22 AAG73837	Human colon cancer

10	1342.5	46.3	673	22	ABB66506	Drosophila melanog
11	1298.5	44.8	287	20	AAY60382	Human normal bladd
12	936.5	32.3	593	21	AA443002	Human ORFX ORF2766
13	936.5	32.3	593	22	AA440333	Human polypeptide
14	895.5	30.9	524	22	AA440332	Human polypeptide
15	880.5	30.4	586	22	AA440332	Human polypeptide
16	789.	27.2	161	23	ABB77790	Amino acid sequenc
17	692	23.9	197	22	AA440332	Human colon cancer
18	666	23.0	216	21	AA440332	Human colon cancer
19	624.5	21.5	264	20	AA440332	Fragment of human
20	532.5	18.4	140	21	AA440332	Human secreted pro
21	411	14.2	151	22	ABB12296	Human APOBEC-1 sc1
22	411	14.2	151	22	AA440733	Human polypeptide
23	395.5	13.6	182	22	AA442118	Human polypeptide
24	395.5	13.6	182	22	AA442119	Human polypeptide
25	330	11.4	124	22	ABG47622	Human liver peptid
26	330	11.4	124	22	ABB27600	Human peptide #251
27	330	11.4	124	22	ABB32770	Peptide #276 encod
28	330	11.4	124	22	ABB18253	Protein #252 encod
29	330	11.4	124	22	AA440332	Human brain expres
30	330	11.4	124	22	AA440332	Human bone marrow
31	330	11.4	124	22	AA440332	Peptide #259 encod
32	330	11.4	124	22	AA440332	Peptide #269 encod
33	330	11.4	124	22	AA440332	Peptide #251 encod
34	330	11.4	124	22	AA440332	Human peptide enco
35	299.5	10.3	652	22	AA440332	Corn poly (A) bind
36	299.5	10.3	652	22	AA440332	Glycine max poly (
37	294.5	10.2	629	23	ABG93292	C. albicans BAX-88
38	273	9.4	55	22	AA440332	Human secreted pro
39	269	9.3	655	22	AA440332	Rice poly (A) bind
40	266	9.2	503	21	AA440332	Arabidopsis thalia
41	266	9.2	515	21	AA440332	Arabidopsis thalia
42	266	9.2	547	21	AA440332	Arabidopsis thalia
43	266	9.2	559	21	AA440332	Arabidopsis thalia
44	266	9.2	618	21	AA440332	Arabidopsis thalia
45	266	9.2	662	21	AA440332	Arabidopsis thalia

ALIGNMENTS

RESULT 1
AAY24418
ID AAY24418 standard; Protein; 545 AA.

XX AAY24418;

AC AAY24418;

XX 21-SEP-1999 (first entry)

XX Human nucleolin-like protein.

DE Human nucleolin-like protein.

DE Human nucleolin-like protein.

XX Human; nucleolin-like protein; HNLP; cancer; melanoma; breast cancer;

XX Human; nucleolin-like protein; HNLP; cancer; melanoma; breast cancer;

XX Human; nucleolin-like protein; HNLP; cancer; melanoma; breast cancer;

XX Human; nucleolin-like protein; HNLP; cancer; melanoma; breast cancer;

XX Human; nucleolin-like protein; HNLP; cancer; melanoma; breast cancer;

XX Human; nucleolin-like protein; HNLP; cancer; melanoma; breast cancer;

XX Human; nucleolin-like protein; HNLP; cancer; melanoma; breast cancer;

XX Human; nucleolin-like protein; HNLP; cancer; melanoma; breast cancer;

XX Human; nucleolin-like protein; HNLP; cancer; melanoma; breast cancer;

XX Human; nucleolin-like protein; HNLP; cancer; melanoma; breast cancer;

XX Human; nucleolin-like protein; HNLP; cancer; melanoma; breast cancer;

XX Human; nucleolin-like protein; HNLP; cancer; melanoma; breast cancer;

PT diagnosis, prevention and treatment of disorders associated with
PT abnormal nucleolin-like protein expression such as cancers,
PT autoimmune disorders and Alzheimer's disease
XX Claim 8; Fig 1; 33pp; English.

XX The present sequence represents human nucleolin-like peptide, designated
CC HNLN. HNLN is the main protein component in the nucleolus of eukaryotic
CC cells and is an essential part of ribosome biosynthesis and also plays
CC an important role in importing proteins to the nucleolus. HNLN may be used
CC in the diagnosis, prevention and treatment of disorders associated with
CC abnormal expression of HNLN. For example, it may be used to treat
CC cancers (e.g. melanoma, breast cancer and prostate cancer), autoimmune
CC disorders (e.g. autoimmune haemolytic anaemia and inflammatory bowel
CC disease) and, in particular, Alzheimer's disease. It may also be used to
CC study the function of the HNLN peptides, the formation (biosynthesis) of
CC ribosomes and the intake of proteins into the nucleolus.

XX SQ Sequence 545 AA;
Query Match 100.0%; Score 2898; DB 20; Length 545;
Best Local Similarity 100.0%; Pred. No. 7.6e-243;
Matches 545; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
Qy 1 MATEHVNGNGTEEPMDTTSVAIHSENFQTLDDAGLPQKVAEKLEIYVAGLVAHSDLD 60
Db 1 MATEHVNGNGTEEPMDTTSVAIHSENFQTLDDAGLPQKVAEKLEIYVAGLVAHSDLD 60
Qy 61 AIEALKEFNEDGALAVLQPKDSLSHVQNKSAFLCGVMKTYRQREKQGTKVADSSK 120
Db 61 AIEALKEFNEDGALAVLQPKDSLSHVQNKSAFLCGVMKTYRQREKQGTKVADSSK 120
Qy 121 EAKIKALLERTGYTLDTVTGQKRGPPPSVYSGQPSVGTETFPVGIKIPDLPEDEL 180
Db 121 EAKIKALLERTGYTLDTVTGQKRGPPPSVYSGQPSVGTETFPVGIKIPDLPEDEL 180
Qy 181 LFEKAGPIWDLRLMMDPLTGLNRYGAFVFTCTKEAAQEAQVAVKLYNNHIRSGKHIG 240
Db 181 LFEKAGPIWDLRLMMDPLTGLNRYGAFVFTCTKEAAQEAQVAVKLYNNHIRSGKHIG 240
Qy 241 ANNLFVGSIPKSKTKQILKEEFKVTGLTDVILYHQPDDKKNRGFCFLEYEDHKTAA 300
Db 241 ANNLFVGSIPKSKTKQILKEEFKVTGLTDVILYHQPDDKKNRGFCFLEYEDHKTAA 300
Qy 301 QARBLMSGKVKVWNGVGTVEWADPIEDDPVMAKVVLVFNLANVTVEIILEKAFSQ 360
Db 301 QARBLMSGKVKVWNGVGTVEWADPIEDDPVMAKVVLVFNLANVTVEIILEKAFSQ 360
Qy 361 FGKLERVKKLKDYAFIHFDERDGAQVAMEENMGKDLGEGNIEIVFAKPPQKPKERKA 420
Db 361 FGKLERVKKLKDYAFIHFDERDGAQVAMEENMGKDLGEGNIEIVFAKPPQKPKERKA 420
Qy 421 QAAKNQMYDDYYGPPHMPPTGRGRGGGYYGPPDYGYEDYDYDYGYDHYNRRGG 480
Db 421 QAAKNQMYDDYYGPPHMPPTGRGRGGGYYGPPDYGYEDYDYDYGYDHYNRRGG 480
Qy 481 YEDPYGYEDFQVARGRGGRGARGAAPSRCRGAAPRGRAGYSQRCGPGSARGVRAG 540
Db 481 YEDPYGYEDFQVARGRGGRGARGAAPSRCRGAAPRGRAGYSQRCGPGSARGVRAG 540
Qy 541 GRGRS 545
Db 541 GRGRS 545

RESULT 2
ID ABG32846 standard; Protein; 545 AA.
XX ABG32846;
AC ABG32846;
XX 18-NOV-2002 (first entry)
XX

DE Human nucleolin-like protein, HNLN.
XX Human; nucleolin-like protein; HNLN; autoimmune disorder;
KW acquired immunodeficiency syndrome; AIDS; Addison's disease; allergy;
KW asthma; atherosclerosis; multiple sclerosis; rheumatoid arthritis;
KW osteoporosis; viral infection; bacterial infection; fungal infection;
KW parasitic infection; protozoal infection; helminthic infection; cancer;
KW Alzheimer's disease; systemic sclerosis; graft-versus-host disease;
KW systemic lupus erythematosus; INCYTE 2809795.
XX Homo sapiens.
OS US2002098566-A1.
XX 25-JUL-2002.
PD 15-OCT-2001; 2001US-0978242.
XX 12-DEC-1997; 97US-0990114.
PR 01-FEB-1999; 99US-0241333.
XX (INCY-) INCYTE PHARM INC.
XX Bandman O, Yue H, Corley NC, Shah P;
PI WPI; 2002-690482/74.
XX N-PSDB; ABS52999.
DR Novel human nucleolin-like polypeptide, useful in diagnosis, prevention
XX and treatment of cancer, Alzheimer's disease and autoimmune disorder
PT such as AIDS, Addison's disease, allergy, asthma, and atherosclerosis -
PT Claim 1; Fig 1; 37pp; English.
XX The invention relates to an isolated human nucleolin-like polypeptide
XX (HNLN) (S1), a polypeptide comprising a naturally occurring sequence
CC having at least 90% identity to S1, or a biologically active or
CC immunogenic fragment of S1 and the HNLN encoding nucleic acid. HNLN is
CC useful for screening a compound for effectiveness as an agonist or
CC antagonist, for screening a compound that specifically binds HNLN or
CC modulates the activity of HNLN, and for preparing a polyclonal or
CC monoclonal antibody by hybridoma technology. HNLN nucleic acid is useful
CC for screening a compound for effectiveness in altering expression of a
CC target polynucleotide comprising HNLN nucleic acid and HNLN probes are
CC useful for assessing toxicity of a test compound. Anti-HNLN antibody is
CC useful in a diagnostic test for a condition or a disease associated with
CC the expression of HNLN in a biological sample, for detecting HNLN in a
CC sample, and for purifying HNLN from a sample. HNLN ant/agonists are
CC useful for treating a disease or condition associated with decreased or
CC increased expression of functional HNLN. The antibody is useful
CC for diagnosing a condition or disease associated with the expression of
CC HNLN in a subject. A HNLN nucleic acid microarray is useful for
CC generating a transcript image of a sample which contains polynucleotides.
CC HNLN and its nucleic acid are useful for diagnosing, treating and
CC preventing an autoimmune disorder (e.g. acquired immunodeficiency
CC syndrome (AIDS), Addison's disease, allergy, asthma, atherosclerosis,
CC multiple sclerosis, rheumatoid arthritis) osteoporosis, cancer, viral,
CC fungal, parasitic, protozoal, helminthic infections, cancer, Alzheimer's
CC disease, systemic sclerosis, graft-versus-host disease and
CC systemic lupus erythematosus (many more diseases are listed in the
CC specification). The present sequence is the human HNLN protein encoded by
CC a cDNA from INCYTE clone 2809795.
XX Sequence 545 AA;
SQ Query Match 100.0%; Score 2898; DB 23; Length 545;
Best Local Similarity 100.0%; Pred. No. 7.6e-243;
Matches 545; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
Qy 1 MATEHVNGNGTEEPMDTTSVAIHSENFQTLDDAGLPQKVAEKLEIYVAGLVAHSDLD 60
Db 1 MATEHVNGNGTEEPMDTTSVAIHSENFQTLDDAGLPQKVAEKLEIYVAGLVAHSDLD 60

```

Best Local Similarity 100.0%; Pred. No. 6.2e-239;
Matches 545; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MATEHVNNGTREPMDTTSAVIHSENFOTLLDAGLPQKVAEKLDEIYVAGLVAHSDLDER 60
Db 1 MATEHVNNGTREPMDTTSAVIHSENFOTLLDAGLPQKVAEKLDEIYVAGLVAHSDLDER 60
QY 61 ATEALKEFNEDGALAVLQFQKDSLSHVQNSAFVCTKEAAQAEAVKLYNNHEIRSGKHIGVCISV 120
Db 61 ATEALKEFNEDGALAVLQFQKDSLSHVQNSAFVCTKEAAQAEAVKLYNNHEIRSGKHIGVCISV 120
QY 121 EAKIKALLERTGYTLDTVTGQKRYGPPPPDSVYSGQPSVGTETIFVGIKIPRDLFEDELVP 180
Db 121 EAKIKALLERTGYTLDTVTGQKRYGPPPPDSVYSGQPSVGTETIFVGIKIPRDLFEDELVP 180
QY 181 LFEKAGPIWDLRLMMDPLTGLNRGYAFVTFCTKEAAQAEAVKLYNNHEIRSGKHIGVCISV 240
Db 181 LFEKAGPIWDLRLMMDPLTGLNRGYAFVTFCTKEAAQAEAVKLYNNHEIRSGKHIGVCISV 240
QY 241 ANRLFVGSIPKSKTKEQILEEFSKVTEGLTDVILYHQPDDKKNGRCFLEYEDHKTAA 300
Db 241 ANRLFVGSIPKSKTKEQILEEFSKVTEGLTDVILYHQPDDKKNGRCFLEYEDHKTAA 300
QY 301 QARRRLMSGKVKVGNVGTVEWADPIEDDPDEVMAKVLFVRNLANTVTTEEILEKAFSQ 360
Db 301 QARRRLMSGKVKVGNVGTVEWADPIEDDPDEVMAKVLFVRNLANTVTTEEILEKAFSQ 360
QY 361 FGKLERVKLKDYAFIHFDERDGA VKAMEEWMNGKLEGENIEIVFAKPPDQKREKKAOR 420
Db 361 FGKLERVKLKDYAFIHFDERDGA VKAMEEWMNGKLEGENIEIVFAKPPDQKREKKAOR 420
QY 421 QAAKNQMYDDYYGPPHMPPTTRGRGRGGYGYPPDYGYEDYDYGYDHYNRYGG 480
Db 421 QAAKNQMYDDYYGPPHMPPTTRGRGRGGYGYPPDYGYEDYDYGYDHYNRYGG 480
QY 481 YEDPYGYEDFQVARGRGGRGARGAAPSRGGAAPPRGRAGYSGRGPGSARGVRAGKR 540
Db 481 YEDPYGYEDFQVARGRGGRGARGAAPSRGGAAPPRGRAGYSGRGPGSARGVRAGKR 540
QY 541 GRGRS 545
Db 541 GRGRS 545

```

RESULT 2

```

US-09-241-333-1
; Sequence 1, Application US/09241333
; Patent No. 6313266
; GENERAL INFORMATION:
; APPLICANT: Bandman, Olga
; APPLICANT: Yue, Henry
; APPLICANT: Corley, Neil C.
; APPLICANT: Shah, Purvi
; TITLE OF INVENTION: HUMAN NUCLEOLIN-LIKE PROTEIN
; NUMBER OF SEQUENCES: 3
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Incyte Pharmaceuticals, Inc.
; STREET: 3174 Porter Drive
; CITY: Palo Alto
; STATE: CA
; COUNTRY: USA
; ZIP: 94304
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette
; OPERATING SYSTEM: DOS
; SOFTWARE: FASTSEQ for Windows Version 2.0
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/241,333
; FILING DATE:
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/990,114

```

```

; PILING DATE:
; ATTORNEY/AGENT INFORMATION:
; NAME: Billings, Lucy J.
; REGISTRATION NUMBER: 36,749
; REFERENCE/DOCKET NUMBER: PF-0451 US
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 650-855-0555
; TELEFAX: 650-845-4166
; TELEX:
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 545 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; IMMEDIATE SOURCE:
; LIBRARY: LYNNNOT05
; CLONE: 2809795
; US-09-241-333-1

```

```

Query Match 100.0%; Score 2898; DB 4; Length 545;
Best Local Similarity 100.0%; Pred. No. 6.2e-239;
Matches 545; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MATEHVNNGTREPMDTTSAVIHSENFOTLLDAGLPQKVAEKLDEIYVAGLVAHSDLDER 60
Db 1 MATEHVNNGTREPMDTTSAVIHSENFOTLLDAGLPQKVAEKLDEIYVAGLVAHSDLDER 60
QY 61 ATEALKEFNEDGALAVLQFQKDSLSHVQNSAFVCTKEAAQAEAVKLYNNHEIRSGKHIGVCISV 120
Db 61 ATEALKEFNEDGALAVLQFQKDSLSHVQNSAFVCTKEAAQAEAVKLYNNHEIRSGKHIGVCISV 120
QY 121 EAKIKALLERTGYTLDTVTGQKRYGPPPPDSVYSGQPSVGTETIFVGIKIPRDLFEDELVP 180
Db 121 EAKIKALLERTGYTLDTVTGQKRYGPPPPDSVYSGQPSVGTETIFVGIKIPRDLFEDELVP 180
QY 181 LFEKAGPIWDLRLMMDPLTGLNRGYAFVTFCTKEAAQAEAVKLYNNHEIRSGKHIGVCISV 240
Db 181 LFEKAGPIWDLRLMMDPLTGLNRGYAFVTFCTKEAAQAEAVKLYNNHEIRSGKHIGVCISV 240
QY 241 ANRLFVGSIPKSKTKEQILEEFSKVTEGLTDVILYHQPDDKKNGRCFLEYEDHKTAA 300
Db 241 ANRLFVGSIPKSKTKEQILEEFSKVTEGLTDVILYHQPDDKKNGRCFLEYEDHKTAA 300
QY 301 QARRRLMSGKVKVGNVGTVEWADPIEDDPDEVMAKVLFVRNLANTVTTEEILEKAFSQ 360
Db 301 QARRRLMSGKVKVGNVGTVEWADPIEDDPDEVMAKVLFVRNLANTVTTEEILEKAFSQ 360
QY 361 FGKLERVKLKDYAFIHFDERDGA VKAMEEWMNGKLEGENIEIVFAKPPDQKREKKAOR 420
Db 361 FGKLERVKLKDYAFIHFDERDGA VKAMEEWMNGKLEGENIEIVFAKPPDQKREKKAOR 420
QY 421 QAAKNQMYDDYYGPPHMPPTTRGRGRGGYGYPPDYGYEDYDYGYDHYNRYGG 480
Db 421 QAAKNQMYDDYYGPPHMPPTTRGRGRGGYGYPPDYGYEDYDYGYDHYNRYGG 480
QY 481 YEDPYGYEDFQVARGRGGRGARGAAPSRGGAAPPRGRAGYSGRGPGSARGVRAGKR 540
Db 481 YEDPYGYEDFQVARGRGGRGARGAAPSRGGAAPPRGRAGYSGRGPGSARGVRAGKR 540
QY 541 GRGRS 545
Db 541 GRGRS 545

```

RESULT 3

```

US-09-347-833-2
; Sequence 2, Application US/09347833
; Patent No. 6294658
; GENERAL INFORMATION:
; APPLICANT: Famodu, Layo O.
; APPLICANT: Odell, Joan T.
; TITLE OF INVENTION: Factors Involved in Gene Expression

```

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: January 28, 2004, 09:23:45 ; Search time 21 Seconds
(without alignments)
1098.068 Million cell updates/sec

Title: US-09-978-242-1
Perfect score: 2898
Sequence: 1 MATEHVNGNGTEPMDTTSA.....RCGPGSARGVRACKRGGRGS 545

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 328717 seqs, 42310858 residues

Total number of hits satisfying chosen parameters: 328717

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Issued Patents AA:*
1: /cgn2_6/ptodata/1/iaa/5A COMB.pcp.*
2: /cgn2_6/ptodata/1/iaa/5B COMB.pcp.*
3: /cgn2_6/ptodata/1/iaa/6A COMB.pcp.*
4: /cgn2_6/ptodata/1/iaa/6B COMB.pcp.*
5: /cgn2_6/ptodata/1/iaa/PCTUS COMB.pcp.*
6: /cgn2_6/ptodata/1/iaa/backfiles1.pcp.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match %	Length	DB ID	Description
1	2898	100.0	545	2	US-08-990-114-1
2	2898	100.0	545	4	US-09-241-333-1
3	299.5	10.3	652	3	US-09-347-833-2
4	269	9.3	655	3	US-09-347-833-4
5	258.5	8.9	714	2	US-08-990-114-3
6	258.5	8.9	714	4	US-09-241-333-3
7	223.5	7.7	336	1	US-07-667-276A-8
8	209	7.2	359	1	US-07-881-075-2
9	209	7.2	359	1	US-08-120-827-2
10	209	7.2	359	1	US-08-478-675-2
11	206.5	7.1	380	1	US-07-881-075-51
12	206.5	7.1	380	1	US-08-120-827-51
13	206.5	7.1	380	1	US-08-478-675-51
14	205.5	7.1	688	3	US-08-973-273-5
15	204.5	7.1	688	3	US-08-973-273-26
16	204.5	7.1	747	3	US-08-973-273-3
17	197.5	6.8	341	2	US-08-538-711A-8
18	197.5	6.8	341	3	US-08-725-027-8
19	197.5	6.8	341	4	US-09-542-552-8
20	197.5	6.8	353	2	US-08-538-711A-7
21	197.5	6.8	353	3	US-08-725-027-7
22	197.5	6.8	353	4	US-09-542-552-7
23	197.5	6.8	444	1	US-07-881-075-3
24	197.5	6.8	444	1	US-08-120-827-3
25	197.5	6.8	444	1	US-08-478-675-3
26	194.5	6.7	223	1	US-07-667-276A-10
27	189	6.5	428	3	US-09-347-833-6

28	184	6.3	437	2	US-08-935-450-8	Sequence 8, Appli
29	181.5	6.3	323	1	US-07-667-276A-7	Sequence 7, Appli
30	181.5	6.3	485	1	US-07-881-075-1	Sequence 1, Appli
31	181.5	6.3	485	1	US-08-120-827-1	Sequence 1, Appli
32	181.5	6.3	485	1	US-08-478-675-1	Sequence 1, Appli
33	181	6.2	759	1	US-08-676-967-1	Sequence 1, Appli
34	181	6.2	759	1	US-08-676-974-1	Sequence 1, Appli
35	181	6.2	759	2	US-09-098-487-1	Sequence 1, Appli
36	179.5	6.2	284	1	US-08-390-858B-36	Sequence 36, Appli
37	177	6.1	656	2	US-08-343-443B-2	Sequence 2, Appli
38	177	6.1	656	3	US-09-214-564A-4	Sequence 4, Appli
39	169.5	5.8	471	3	US-08-866-928B-1	Sequence 1, Appli
40	169.5	5.8	471	4	US-09-685-836-1	Sequence 1, Appli
41	169	5.7	287	1	US-08-146-421-5	Sequence 5, Appli
42	166.5	5.7	414	1	US-07-667-276A-4	Sequence 4, Appli
43	158.5	5.5	217	1	US-08-390-858B-9	Sequence 9, Appli
44	147.5	5.1	162	4	US-09-575-574-4	Sequence 4, Appli
45	145.5	5.0	425	4	US-09-252-991A-17013	Sequence 17013, A

ALIGNMENTS

RESULT 1
US-08-990-114-1
; Sequence 1, Application US/08990114
; Patent No. 5932475
; GENERAL INFORMATION:
; APPLICANT: Bandman, Olga
; APPLICANT: Yue, Henry
; APPLICANT: Corley, Neil C.
; APPLICANT: Shah, Purvi
; TITLE OF INVENTION: HUMAN NUCLEOLIN-LIKE PROTEIN
; NUMBER OF SEQUENCES: 3
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Incyte Pharmaceuticals, Inc.
; STREET: 3174 Porter Drive
; CITY: Palo Alto
; STATE: CA
; COUNTRY: USA
; ZIP: 94304
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette
; COMPUTER: IBM Compatible
; OPERATING SYSTEM: DOS
; SOFTWARE: FastSeq for Windows Version 2.0
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/990,114
; FILING DATE: Herewith
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER:
; FILING DATE:
; ATTORNEY/AGENT INFORMATION:
; NAME: Billings, Lucy J.
; REGISTRATION NUMBER: 36,749
; REFERENCE/DOCKET NUMBER: PF-0451 US
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 650-855-0555
; TELEFAX: 650-845-4166
; TELEX:
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 545 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; IMMEDIATE SOURCE:
; LIBRARY: TLYNOT05
; CLONE: 2809795
; US-08-990-114-1

Query Match 100.0%; Score 2898; DB 2; Length 545;

TELEFAX: 650-845-4166
TELEX: <Unknown>
INFORMATION FOR SEQ ID NO: 1:
SEQUENCE CHARACTERISTICS:
LENGTH: 545 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
IMMEDIATE SOURCE:
LIBRARY: TLYNOT05
CLONE: 2809795
SEQUENCE DESCRIPTION: SEQ ID NO: 1:
US-09-978-242-1

Query Match 100.0%; Score 2898; DB 9; Length 545;
Best Local Similarity 100.0%; Pred. No. 8.3e-239; Indels 0; Gaps 0;
Matches 545; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 MATEHVNGNGTEPEMDTTSAVIHSENFQTLDDAGLPQKVAEKLDEIYVAGLVASHDLDER 60
Db 1 MATEHVNGNGTEPEMDTTSAVIHSENFQTLDDAGLPQKVAEKLDEIYVAGLVASHDLDER 60

Qy 61 AIEALKEFNEDGALAVLQOFKDSLSHVQNKSAFLCGVMKTYRQREKQGTKVADSSKGP 120
Db 61 AIEALKEFNEDGALAVLQOFKDSLSHVQNKSAFLCGVMKTYRQREKQGTKVADSSKGP 120

Qy 121 EAKIKALLERTGYTLDTVTGQRYGGPPDSVYSGQSPSVGTEIFVGGKIPRDLFEDELVP 180
Db 121 EAKIKALLERTGYTLDTVTGQRYGGPPDSVYSGQSPSVGTEIFVGGKIPRDLFEDELVP 180

Qy 181 LFEKAGPIWDLRLMMDPLTGLNRGYAFVTFCTKEAAQAEAVKLYNNHEIRSGKHIGVCISV 240
Db 181 LFEKAGPIWDLRLMMDPLTGLNRGYAFVTFCTKEAAQAEAVKLYNNHEIRSGKHIGVCISV 240

Qy 241 ANNRLFVGSIPKSKTKQIILEEFSKVTEGLTDVILYHQDDKKNGRCFLEYEDHKTAA 300
Db 241 ANNRLFVGSIPKSKTKQIILEEFSKVTEGLTDVILYHQDDKKNGRCFLEYEDHKTAA 300

Qy 301 QARRRLMSGKVKVNGVGTVEWADPIEDDPPEVMAKVLFVRNLANTVTTEEILEKAFSQ 360
Db 301 QARRRLMSGKVKVNGVGTVEWADPIEDDPPEVMAKVLFVRNLANTVTTEEILEKAFSQ 360

Qy 361 FGKLERVKKLDYAFIHFDERDGAVKAMEENMGKLEGENIEIVFAKPPDQKREKKAQR 420
Db 361 FGKLERVKKLDYAFIHFDERDGAVKAMEENMGKLEGENIEIVFAKPPDQKREKKAQR 420

Qy 421 QAAKNQMYDDYYGYGPPHMPPTTRGRGRGGYGYPPDYGYEDYDYGYDHYNRYGG 480
Db 421 QAAKNQMYDDYYGYGPPHMPPTTRGRGRGGYGYPPDYGYEDYDYGYDHYNRYGG 480

Qy 481 YEDPYGYEDFQVARGRGARGAAPSRCGAAPPRGRAGYSORGGPGSARGVRAGKR 540
Db 481 YEDPYGYEDFQVARGRGARGAAPSRCGAAPPRGRAGYSORGGPGSARGVRAGKR 540

Qy 541 GRGRS 545
Db 541 GRGRS 545

RESULT 2
US-09-925-300-1415
Sequence 1415, Application US/09925300
Patent No. US20020151681A1
GENERAL INFORMATION:
APPLICANT: Craig Rosen,
APPLICANT: Steve Ruben
TITLE OF INVENTION: Nucleic Acids, Proteins and Antibodies
FILE REFERENCE: PA101
CURRENT APPLICATION NUMBER: US/09/925,300
PRIOR FILING DATE: 2001-08-10
PRIOR APPLICATION NUMBER: PCT/US00/05988
PRIOR FILING DATE: 2000-03-08
PRIOR APPLICATION NUMBER: 60/124,270

PRIOR FILING DATE: 1999-03-12
NUMBER OF SEQ ID NOS: 1890
SOFTWARE: PatentIn Ver. 2.0
SEQ ID NO 1415
LENGTH: 579
TYPE: PRT
ORGANISM: Homo sapiens
US-09-925-300-1415

Query Match 98.8%; Score 2864; DB 10; Length 579;
Best Local Similarity 98.2%; Pred. No. 7.3e-236; Indels 6; Gaps 1;
Matches 540; Conservative 2; Mismatches 2; Indels 6; Gaps 1;

Qy 1 MATEHVNGNGTEPEMDTTSAVIHSENFQTLDDAGLPQKVAEKLDEIYVAGLVASHDLDER 60
Db 1 MATEHVNGNGTEPEMDTTSAVIHSENFQTLDDAGLPQKVAEKLDEIYVAGLVASHDLDER 77

Qy 61 AIEALKEFNEDGALAVLQOFKDSLSHVQNKSAFLCGVMKTYRQREKQGTKVADSSKGP 120
Db 61 AIEALKEFNEDGALAVLQOFKDSLSHVQNKSAFLCGVMKTYRQREKQGTKVADSSKGP 137

Qy 121 EAKIKALLERTGYTLDTVTGQRYGGPPDSVYSGQSPSVGTEIFVGGKIPRDLFEDELVP 180
Db 121 EAKIKALLERTGYTLDTVTGQRYGGPPDSVYSGQSPSVGTEIFVGGKIPRDLFEDELVP 197

Qy 181 LFEKAGPIWDLRLMMDPLTGLNRGYAFVTFCTKEAAQAEAVKLYNNHEIRSGKHIGVCISV 240
Db 181 LFEKAGPIWDLRLMMDPLTGLNRGYAFVTFCTKEAAQAEAVKLYNNHEIRSGKHIGVCISV 257

Qy 241 ANNRLFVGSIPKSKTKQIILEEFSKVTEGLTDVILYHQDDKKNGRCFLEYEDHKTAA 300
Db 241 ANNRLFVGSIPKSKTKQIILEEFSKVTEGLTDVILYHQDDKKNGRCFLEYEDHKTAA 317

Qy 301 QARRRLMSGKVKVNGVGTVEWADPIEDDPPEVMAKVLFVRNLANTVTTEEILEKAFSQ 360
Db 301 QARRRLMSGKVKVNGVGTVEWADPIEDDPPEVMAKVLFVRNLANTVTTEEILEKAFSQ 377

Qy 361 FGKLERVKKLDYAFIHFDERDGAVKAMEENMGKLEGENIEIVFAKPPDQKREKKAQR 420
Db 361 FGKLERVKKLDYAFIHFDERDGAVKAMEENMGKLEGENIEIVFAKPPDQKREKKAQR 437

Qy 421 QAAKNQMYDDYYGYGPPHMPPTTRGRGRGGYGYPPDYGYEDYDYGYDHYNRYGG 480
Db 421 QAAKNQMYDDYYGYGPPHMPPTTRGRGRGGYGYPPDYGYEDYDYGYDHYNRYGG 497

Qy 481 YEDPYGYEDFQVARGRGARGAAPSRCGAAPPRGRAGYSORGGPGSARGVRAGK - 539
Db 481 YEDPYGYEDFQVARGRGARGAAPSRCGAAPPRGRAGYSORGGPGSARGVRAGK 557

Qy 540 -----RGRGR 544
Db 540 GAOQQRGRGQ 567

RESULT 3
US-10-106-698-5319
Sequence 5319, Application US/10106698
Publication No. US20030109690A1
GENERAL INFORMATION:
APPLICANT: Ruben et al.
TITLE OF INVENTION: Colon and Colon Cancer Associated Polynucleotides and Polypeptides
FILE REFERENCE: PA005P1
CURRENT APPLICATION NUMBER: US/10/106,698
CURRENT FILING DATE: 2002-03-27
PRIOR APPLICATION NUMBER: PCT/US00/26524
PRIOR FILING DATE: 2000-09-28
PRIOR APPLICATION NUMBER: US 60/157,137
PRIOR FILING DATE: 1999-09-29
PRIOR APPLICATION NUMBER: US 60/163,280
PRIOR FILING DATE: 1999-11-03
NUMBER OF SEQ ID NOS: 8564
SOFTWARE: PatentIn Ver. 3.0
SEQ ID NO 5319

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: January 28, 2004, 09:26:11 ; Search time 38 Seconds
(without alignments)

2965.039 Million cell updates/sec

Title: US-09-978-242-1

Perfect score: 2898

Sequence: 1 MATEHVNGNGTEPMDTTSA.....RGPGSGARGVRAGKRGGRS 545

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 777136 seqs, 206736638 residues

Total number of hits satisfying chosen parameters: 777136

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Listing first 45 summaries

Database : Published Applications AA:*

1: /cgn2_6/ptodata/2/pubpaa/US07_PUBCOMB.pep.*
2: /cgn2_6/ptodata/2/pubpaa/PCT_NEW_PUB.pep.*
3: /cgn2_6/ptodata/2/pubpaa/US06_NEW_PUB.pep.*
4: /cgn2_6/ptodata/2/pubpaa/US06_PUBCOMB.pep.*
5: /cgn2_6/ptodata/2/pubpaa/US07_NEW_PUB.pep.*
6: /cgn2_6/ptodata/2/pubpaa/PCTUS_PUBCOMB.pep.*
7: /cgn2_6/ptodata/2/pubpaa/US08_NEW_PUB.pep.*
8: /cgn2_6/ptodata/2/pubpaa/US08_PUBCOMB.pep.*
9: /cgn2_6/ptodata/2/pubpaa/US09A_PUBCOMB.pep.*
10: /cgn2_6/ptodata/2/pubpaa/US09B_PUBCOMB.pep.*
11: /cgn2_6/ptodata/2/pubpaa/US09C_PUBCOMB.pep.*
12: /cgn2_6/ptodata/2/pubpaa/US09_NEW_PUB.pep.*
13: /cgn2_6/ptodata/2/pubpaa/US10A_PUBCOMB.pep.*
14: /cgn2_6/ptodata/2/pubpaa/US10B_PUBCOMB.pep.*
15: /cgn2_6/ptodata/2/pubpaa/US10C_PUBCOMB.pep.*
16: /cgn2_6/ptodata/2/pubpaa/US10_NEW_PUB.pep.*
17: /cgn2_6/ptodata/2/pubpaa/US60_NEW_PUB.pep.*
18: /cgn2_6/ptodata/2/pubpaa/US60_PUBCOMB.pep.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Query Match	Score	Length	ID	Description
1	2898	100.0	545	9	US-09-978-242-1
2	2864	98.8	579	10	US-09-925-300-1415
3	2863	98.8	704	15	US-10-106-698-5319
4	2860	98.7	561	10	US-09-821-687-4
5	2355	81.3	633	10	US-09-821-687-10
6	2355	81.3	633	11	US-09-374-046A-128
7	1647.5	56.8	397	15	US-10-106-698-4611
8	1342.5	46.3	673	12	US-09-949-029-100
9	1075	37.1	335	10	US-09-821-687-11
10	789	27.2	161	10	US-09-821-687-2
11	692	23.9	197	15	US-10-106-698-6457
12	666	23.0	216	9	US-09-925-299-841
13	666	23.0	216	11	US-09-925-299-841
14	330	11.4	124	9	US-09-864-761-33551
15	266	9.2	662	12	US-10-338-777-42

16	258.5	8.9	714	9	US-09-978-242-3
17	249.5	8.6	402	14	US-10-062-254-258
18	249.5	8.6	687	12	US-10-104-047-2651
19	244	8.4	707	12	US-10-384-569-3
20	236.5	8.2	406	14	US-10-062-254-238
21	234.5	8.1	409	14	US-10-062-254-276
22	233	8.0	314	9	US-09-849-967A-3
23	233	8.0	320	12	US-10-341-434-59
24	232	8.0	428	14	US-10-062-254-250
25	228.5	7.9	430	14	US-10-062-254-252
26	227	7.8	332	12	US-10-108-260A-4506
27	226.5	7.8	633	12	US-10-341-434-49
28	225	7.8	301	9	US-09-799-777-64
29	221.5	7.6	432	14	US-10-062-254-242
30	217	7.5	397	14	US-10-062-254-268
31	216	7.5	378	9	US-09-849-967A-2
32	214.5	7.4	436	14	US-10-062-254-248
33	209	7.2	359	14	US-10-025-367-23
34	208	7.2	416	14	US-10-062-254-264
35	207.5	7.2	350	14	US-10-062-254-260
36	206.5	7.1	380	14	US-10-025-367-27
37	206.5	7.1	420	14	US-10-062-254-254
38	205	7.1	117	9	US-09-764-887-194
39	205	7.1	117	15	US-10-073-961-194
40	202	7.0	494	9	US-09-833-790-234
41	201	6.9	356	14	US-10-062-254-262
42	197.5	6.8	353	11	US-09-935-642-14
43	194.5	6.7	359	14	US-10-025-367-25
44	193.5	6.7	316	14	US-10-062-254-266
45	192.5	6.6	307	14	US-10-062-254-244

ALIGNMENTS

RESULT 1

US-09-978-242-1
; Sequence 1, Application US/09978242
; Patent No. US20020098566A1
; GENERAL INFORMATION:
; APPLICANT: Bandman, Olga
; Yue, Henry
; Corley, Neil C.
; Shah, Purvi
; TITLE OF INVENTION: HUMAN NUCLEOLIN-LIKE PROTEIN
; NUMBER OF SEQUENCES: 3
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Incyte Pharmaceuticals, Inc.
; STREET: 3174 Porter Drive
; CITY: Palo Alto
; STATE: CA
; COUNTRY: USA
; ZIP: 94304
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette
; COMPUTER: IBM Compatible
; OPERATING SYSTEM: DOS
; SOFTWARE: FastSeq for Windows Version 2.0
; CURRENT APPLICATION NUMBER: US/09/978,242
; APPLICATION NUMBER: 15-Oct-2001
; FILING DATE: 15-Oct-2001
; CLASSIFICATION: <Unknown>
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 09/241,333
; FILING DATE: <Unknown>
; APPLICATION NUMBER: 08/990,114
; FILING DATE: <Unknown>
; ATTORNEY/AGENT INFORMATION:
; NAME: Billings, Lucy J.
; REGISTRATION NUMBER: 36,749
; REFERENCE/DOCKET NUMBER: PF-0451 US
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 650-855-0555

Qy 474 YHNYRGYEDPYGYED-FQVARGRGARGA-APSRGRGAAPPRGRAGYSQRGGP-G 530
 Db |||||
 Qy 479 YHNYRGYEDPYGYDGYAVRG-GRGGRGAPPPRGRGAPPRGRAGYSQRGAPLG 537
 Db |||||
 Qy 531 SARVGRAGK-----RGRG 543
 Db |||||
 Qy 538 PPRGSRGGRGGAQQQRGRG 557
 Db |||||

RESULT 2
 T01563
 Hypothetical protein A_TW018A10.14 - Arabidopsis thaliana
 C:Species: Arabidopsis thaliana (mouse-ear cress)
 C:Date: 19-Feb-1999 #sequence_revision 19-Feb-1999 #text_change 24-Mar-1999
 C:Accession: T01563
 R:Dempsey, S.; Harper, M.
 submitted to the EMBL Data Library, July 1997
 A:Description: The sequence of A. thaliana TW018A10.
 A:Reference number: Z14348
 A:Accession: T01563
 A:Status: translated from GB/EMBL/DBJ
 A:Molecule type: DNA
 A:Residues: 1-521 <DEM>
 A:Cross-references: EMBL:AF013294; NID:G2252848; PID:G2252863
 A:Experimental source: cultivar Columbia
 C:Genetics:
 A:Map position: 4
 A:Introns: 158/3; 209/3; 256/3; 316/3; 384/1; 444/3; 473/2
 A:Note: A_TW018A10.14

Query Match 17.1%; Score 495.5; DB 2; Length 521;
 Best Local Similarity 26.2%; Pred. No. 3.9e-24;
 Matches 149; Conservative 97; Mismatches 178; Indels 145; Gaps 17;

Qy 53 AHSÜ-DERATEALKERFEDGALLAVLQFQKDSLSHYONKSAFLCGVWKTQREKQGTK 111
 Db |||||
 Qy 17 SYSEMEDEVEEVEEYEE-----EEEDDDDDVGNQNA-----EEREVED 58
 Db |||||
 Qy 112 VADSSKGDEAKIKALLERTGYTLDTTQKRGKPP-----DSVYSG--QQPSVGTET 164
 Db |||||
 Qy 59 YGDTKGDMEVQBEIEADDDNHDIETADDERPPSPIDDEKYSKSHLLSLPPHSEV 118
 Db |||||
 Qy 165 FVGKIPRDLFDELVLPEKAGPIWD-----LRLMMDPLTGLNRYGAFVTF 210
 Db |||||
 Qy 119 FIGGLPRDVGEDLRLCEIGEIFEVRTAIFVFDHILFVRLMKDRDSGSKGYAFVAF 178
 Db |||||
 Qy 211 CTKEAAQAVKLYNNHEIR-----SGKHIGVCISVANNRFLVGSIPKSKTKBQI 259
 Db |||||
 Qy 179 KTKOVAQKAIHELHSEKFKASTANCSLSLSTKIRCSLSTKIRLFTIGNIPKNWTEDEF 238
 Db |||||
 Qy 260 LEFSKVTETGLTVLYHQDDKKNRGCFLEYEDHKTAAQARRRLMSGKVKVGNVGT 319
 Db |||||
 Qy 239 RKVIEDVGPVENIELIKDPTNTRNRGAFVLYNNACADYSQKMLDSNFKLEGNAPT 298
 Db |||||
 Qy 320 VEWADPIEDDPD-PEYMAKVKVLFVRLNLTNTVEILEKAFSGFKLERVKK-----LKD 372
 Db |||||
 Qy 299 VTWADPKSSPEHSAQAQVAKLYKNIPENTSTQLKELFORHGEVTKIVTPCKGKGRD 358
 Db |||||
 Qy 373 YAFIHFDRDGAVKAMEWNGKDLGENIEIVFAKPPDQKQKRAQQAQAKNQYDDYY 432
 Db |||||
 Qy 359 FGFVHYAERSALKAVKDKTEREYVNGOPLEVVLAKPOAERKHPSS-----Y 405
 Db |||||
 Qy 433 YGPPHPPPTTGRGGRGGYGPDPDYGYDYHYHVRGVEDPYGYEDFQ 492
 Db |||||
 Qy 406 SYGAAPTAPP-----VHPTFGFAAAPYG-----A 431
 Db |||||
 Qy 493 VGARGRGARGAAPSRRGAAP-----PRGRAGY--SORGGP----- 529
 Db |||||
 Qy 432 MGA-GLGITAGSFSQPMIYGRGAMPQMVPMLPLDGRGVGLQPGMPMAAPPQRRR 490
 Db |||||
 Qy 530 -----GSAR-----GVRAGKRGR 542
 Db |||||
 Qy 491 NDRNNGSGGGRONSHEDGNRGRGRYR 519
 Db |||||

RESULT 3

T49019

Probable RNA binding protein - Arabidopsis thaliana

N:Alternate names: protein F3C22.60

C:Species: Arabidopsis thaliana (mouse-ear cress)

C:Date: 02-Jun-2000 #sequence_revision 02-Jun-2000 #text_change 02-Jun-2000

C:Accession: T49019

R:Purnelle, B.; Masuy, D.; Goffeau, A.; Boutry, M.; Mewes, H.W.; Rudd, S.; Lemcke, K.; Mi

submitted to the Protein Sequence Database, April 2000

A:Reference number: Z25013

A:Accession: T49019

A:Status: preliminary

A:Molecule type: DNA

A:Residues: 1-471 <PUR>

A:Cross-references: EMBL:AL353912; GSPDB:GN00061; ATSP:F3C22.60

A:Experimental source: cultivar Columbia; BAC clone F3C22

C:Genetics:

A:Gene: ATSP:F3C22.60

A:Map position: 3

A:Introns: 120/3; 159/3; 206/3; 266/3; 336/1; 395/3; 424/2

Query Match 15.0%; Score 435.5; DB 2; Length 471;

Best Local Similarity 27.6%; Pred. No. 2.4e-20;

Matches 114; Conservative 75; Mismatches 165; Indels 59; Gaps 10;

Qy 158 PSVGTETFGKIPRDLFDELVLPEKAGPIWDLRLMMDPLTGLNRYGAFVTFCTKEAAQ 217

Db |||||

Qy 88 PPHGSEVILGGIPTDATEGDLKGFCSIGEVTEVRIMREKSDGDKGYAFVTRSKDLAA 147

Db |||||

Qy 218 EAVKLYNNHEIRSGKHIGVCISVANNRFLVGSIPKSKTKBQILEEFSKVTEGLTVLYH 277

Db |||||

Qy 148 EADTLNNTDFR-GKIKCSTTQAKHRLFLGNVPRNWMESDIKKAANRIGPGVQIVELPK 206

Db |||||

Qy 278 QPDDKKKRGFCFLEYEDHKTAAQARRRLMSGKVKVGNVGTVEWADPIEDPPPEVMA-K 336

Db |||||

Qy 207 EPQNGMNRGFAFIEYNNHACAEYSKQMSNPSFKLDDNAPTYSWAESRGGGGDSASQ 266

Db |||||

Qy 337 VKVLFVRNLANTVTTEILEKAFSGFKLERV-----KKLKDYAFIHFDRDGAVKAM 388

Db |||||

Qy 267 VKALYITKLPDITQERLKALEFHHGKILKLVIPPAKPKGDSRYGFVHYAERTSVML 326

Db |||||

Qy 389 EEMNGKDLGENIEIVFAKPPDQKQKRAQQAQAKNQYDDYYGPPHPPPTTRGRGR 448

Db |||||

Qy 327 KNTERVEIDGHMLDCTLAKEPADQKTNNTVQNVQSKLPNTY-----PPLL----- 373

Db |||||

Qy 449 GGRGGYGPDPDYGYDYGYDYHYHVRGVEDPYGYEDFQVARGRGARGAAP 508

Db |||||

Qy 374 ----SYGMAPSPFG-----ALGGFGASAYSQPLMHAGHAGHAGSMMPIML 415

Db |||||

Qy 509 SRGR-----GAA-----PPRGRAGYSORGPGGSAR---GVRAGKRGRGRS 545

Db |||||

Qy 416 PDGRIGVILQOPGLAAMPQPPRSPPY---RGGSGSSSSSGSKRSSDNGRGRS 466

Db |||||

RESULT 4

DNZPPA

polyadenylate-binding protein - fission yeast (Schizosaccharomyces pombe)

C:Species: Schizosaccharomyces pombe

C:Date: 14-Feb-1992 #sequence_revision 19-Jan-2001 #text_change 19-Jan-2001

C:Accession: T38950; A39720

R:Skellton, J.; Churcher, C.M.; Barrell, B.G.; Rajandream, M.A.; Wood, V.

submitted to the EMBL Data Library, May 1997

A:Reference number: Z21819

A:Accession: T38950

A:Status: translated from GB/EMBL/DBJ

A:Molecule type: DNA

A:Residues: 1-653 <SKE>

A:Cross-references: EMBL:Z95396; NID:G6090552; PID:G2104439; GSPDB:GN00

A:Experimental source: strain 972h-; cosmid c57A7

R:Burd, C.G.; Matunis, E.L.; Dreyfuss, G.

Mol. Cell. Biol. 11, 3419-3424, 1991

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: January 28, 2004, 09:23:10 ; Search time 20 Seconds
(without alignments)
2620.597 Million cell updates/sec

Title: US-09-978-242-1
Perfect score: 2898
Sequence: 1 MATEHVNGTPEPMDTTSA.....RGSPGARGVRAGKRGGRS 545
Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 283308 seqs, 96168682 residues
Total number of hits satisfying chosen parameters: 283308

Minimum DB seq length: 0
Maximum DB seq length: 2000000000
Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : PIR_76.*
1: pir1.*
2: pir2.*
3: pir3.*
4: pir4.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	2355	81.3	633	2 T02673	heterogeneous nucl
2	495.5	17.1	521	2 T01563	hypothetical prote
3	435.5	15.0	471	2 T49019	probable RNA bindi
4	287.5	9.9	653	1 DNZPPA	polyadenylate-bind
5	286	9.9	651	2 S18874	nucleolin - Africa
6	280	9.7	693	2 JC7925	nucleolin - common
7	279	9.6	671	2 C96534	probable Poly-A Bi
8	278.5	9.6	705	2 S32644	nucleolin - Africa
9	275	9.5	345	1 B41732	heterogeneous nucl
10	274	9.5	651	2 T06979	polyadenylate-bind
11	268.5	9.3	623	2 T07933	polyadenylate-bind
12	266	9.2	662	2 T00497	polyadenylate-bind
13	258.5	8.9	629	2 T05425	polyadenylate-bind
14	258.5	8.9	713	2 A27441	nucleolin - Chinae
15	253	8.7	405	2 H86249	hypothetical prote
16	251.5	8.7	638	2 S37085	polyadenylate-bind
17	251	8.7	448	2 T15542	hypothetical prote
18	249	8.6	353	1 S56750	single stranded D
19	249	8.6	692	2 T21095	hypothetical prote
20	247	8.5	500	2 S55785	nucleolar protein
21	244	8.4	707	2 A35804	nucleolin - human
22	242.5	8.4	636	2 I48718	poly(A) binding pr
23	242	8.4	427	2 T04623	hypothetical prote
24	241.5	8.3	712	2 JH0148	nucleolin - rat
25	239	8.2	320	2 S02061	heterogeneous ribo
26	236.5	8.2	668	2 B96740	hypothetical prote
27	236	8.1	577	1 DNBPA	polyadenylate-bind
28	234.5	8.1	574	2 S30887	polyadenylate-bind
29	233	8.0	320	1 DDRT	helix-destabilizin

30	233	8.0	320	1 A44485	heterogeneous ribo
31	233	8.0	320	2 S04617	heterogeneous ribo
32	233	8.0	707	1 DNMS	nucleolin - mouse
33	232.5	8.0	414	2 JN0866	nucleolar protein
34	232.5	8.0	633	1 DNXLPA	polyadenylate-bind
35	231	8.0	566	2 T21096	hypothetical prote
36	229	7.9	694	1 DNCHNL	nucleolin - chicke
37	227.5	7.9	628	2 S44138	polyadenylate-bind
38	227	7.8	522	2 S52491	polyadenylate-bind
39	226.5	7.8	633	1 DNHUPA	polyadenylate-bind
40	225	7.8	301	2 JW0079	heterogeneous nucl
41	223	7.7	320	2 S30192	heterogeneous ribo
42	222.5	7.7	646	2 T26427	hypothetical prote
43	222	7.7	655	2 T00768	polyadenylate-bind
44	221	7.6	609	2 B84783	probable poly(A) b
45	218.5	7.5	308	2 B47369	RNA-binding protei

ALIGNMENTS

RESULT 1

T02673
heterogeneous nuclear ribonucleoprotein R - human
C;Species: Homo sapiens (man)
C;Date: 24-Mar-1999 #sequence_revision 24-Mar-1999 #text_change 08-Oct-1999
C;Accession: T02673
R;Chan, E.K.L.; Mathison, D.A.; Portman, D.; Dreyfuss, G.; Steiner, G.; Tan, E.M.; Hassfe
Nucleic Acids Res. 26, 439-445, 1998
A;Title: Molecular definition of heterogeneous nuclear ribonucleoprotein R (hnRNP R) usin
A;Reference number: Z14697; MUID:98083170; PMID:9421497
A;Accession: T02673
A;Status: preliminary; translated from GB/EMBL/DDBJ
A;Molecule type: mRNA
A;Residues: 1-633 <CHA>
A;Cross-references: EMBL:AF000364; NID:g2697102; PIDN:AAC39540.1; PID:g2697103

Query Match 81.3%; Score 2355; DB 2; Length 633;
Best Local Similarity 80.2%; Pred. No 5.1e-143;
Matches 449; Conservative 51; Mismatches 40; Indels 20; Gaps 10;

Qy	1	MATEHVNGT-----EPMDDTTSAVHSGENFQTLTDLGLPQKVAEKDLIYVAGLVAHS 55
Db	1	MANQ-VNGNAVQLKEEPEMD-TSSVTHHYKTLIEAGLPQKVAERLDEIFQTLGVAVV 58
Qy	56	DLDERAIEALKEPENEGALAVLQOFKDSLSHVONKSAPFLCGVMKTYRQREKQTKVADS 115
Db	59	DLDERAIDALREFNEEGALSVLQOFKESDLSHVONKSAPFLCGVMKTYRQREKQSKVQES 118
Qy	116	SKGPDSEAKTKALLERTGYTLDTVTGQKYGKGGPPPSVYSGQSPVSGTEIFVGGKIPRDLFE 175
Db	119	TKGPDSEAKTKALLERTGYTLDTVTGQKYGKGGPPPSVYSGQSPVSGTEIFVGGKIPRDLFE 178
Qy	176	DELVPFLFEKAGPIWDLRLMMDPLTGLNRYGAVFTFCTKEAAQEAQVAVKLYNNHIRSKGHG 235
Db	179	DELVPFLFEKAGPIWDLRLMMDPLSGQNRGYAFTTCGKEAAQEAQVAVKLYNNHIRSKGHG 238
Qy	236	VCISVANNRLFVGSIPKSKTKQILIEEFSKVTEGLTDVILYHQDDKKNNRQFCFLEYED 295
Db	239	VCISVANNRLFVGSIPKSKTKQILIEEFSKVTEGLTDVILYHQDDKKNNRQFCFLEYED 298
Qy	296	HKTAQAQRRLMSGKVKVGNVGTVEWADPIEDPPEVMAKVKVLFVRNLANTVTTEEIL 355
Db	299	HKSAQAQRRLMSGKVKVGNVGTVEWADPVEPPEVMAKVKVLFVRNLANTVTTEEIL 358
Qy	356	KAFSGFKLERVKLKDYAFIHFDERDGAVKAMEENKMDLEGENTIEIVFAKPPDKRKE 415
Db	359	KSFSEFGKLERVKLKDYAFVHFDEGAAGAAVKAMDEMNGKEIEGEEIEIVLAKPPDKRKE 418
Qy	416	RIAQQAQAKNQYDDYYYPHPMPPTTGRGR-GRGSGYGPDPYGYVEDYY-DYGYVD 473
Db	419	ROAARQAQRSTAYEDYYYPHPMPPTTGRGRGRGGYGPDPYGYVEDYYDDYYGYD 478

```

FT DOMAIN 412 418 NUCLEAR LOCALIZATION SIGNAL (POTENTIAL) .
FT DOMAIN 165 244 RNA-BINDING (RRM) 1.
FT DOMAIN 246 328 RNA-BINDING (RRM) 2.
FT DOMAIN 341 411 RNA-BINDING (RRM) 3.
FT DOMAIN 447 567 RNA-BINDING (RGG-BOX) .
FT DOMAIN 462 471 3 X 11 AA APPROXIMATE REPEATS OF D-D-Y-Y-
FT REPEAT 471 482 G-Y-D-Y-H-D-Y.
FT REPEAT 482 497 1 (APPROXIMATE) .
FT REPEAT 497 509 2. (APPROXIMATE) .
FT DOMAIN 579 633 GLN/ASN-RICH DOMAIN.
FT SEQUENCE 633 AA; 70943 MW; 088341F6465ED46F CRC64;

Query Match 81.3%; Score 2355; DB 1; Length 633;
Best Local Similarity 80.2%; Pred. No. 8.9e-138;
Matches 449; Conservative 51; Mismatches 40; Indels 20; Gaps 10;

Qy 1 MATEHVNGNGT-----EPPMDTTSVAVIHSNFOTLLDAGLPQKVAEKDEIYVAGLVAHS 55
Db 1 MANQ-VNGNAVQLKEEPEMD-TSSVTHTEHYKTLIEAGLPQKVAERLDBIFQGLVAVY 58
Qy 56 DLDERATEALKERNEDGALVLOQFSDLSHVONKSAFLCGVMKTYRQREKQTKVADS 115
Db 59 DLDERATDALREFNEGALSALVLOQFSDLSHVONKSAFLCGVMKTYRQREKQSKVOES 118
Qy 116 SKGPDEAKIKALLERTGYTLDTVTGQRYKGGPPDPDSVYSGOQPSVGTETIFVGKIPRDLFE 175
Db 119 TKGPDEAKIKALLERTGYTLDTVTGQRYKGGPPDPDSVYSGVQFQIGTEVFVGKIPRDLFE 178
Qy 176 DELVPLFEKAGPIWDLRLMDPLTGLNRGVAFTFTCKEAAQEAVALKLYNNHIEIRSGKHIG 235
Db 179 DELVPLFEKAGPIWDLRLMDPLTGLNRGVAFTFTCKEAAQEAVALKLYNNHIEIRSGKHIG 238
Qy 236 VCISVANNRLPVGSIPIKSKTEQILEEFSKVTGELTDVILYHQDDKKQKRGFCFLEYED 295
Db 239 VCISVANNRLPVGSIPIKSKTEQILEEFSKVTGELTDVILYHQDDKKQKRGFCFLEYED 298
Qy 296 HKTAAQARRRLMSKVKVGNVGTVEWADPTEDPPEVMAKVLFVRNLANTVTTEILE 355
Db 299 HKSAQAARRRLMSKVKVGNVGTVEWADPVEEDPPEVMAKVLFVRNLANTVTTEILE 358
Qy 356 KAFSQFGKLERVKKLKDYAFIHFDERDGAVKAMEMNGKOLEGENIEIVFAKPPDKRKE 415
Db 359 KSFSEFGKLERVKKLKDYAFVHPEDRGAALKAMDMNGKIEEGEEIEIVLAKPPDKRKE 418
Qy 416 RKAQRQAANKQMDYDYIYGGPHPPPTPRGRGR-GRRGGYGYPPDYGYEDYDY-DYGYD 473
Db 419 RQAARQASRSTAYEDYIYHPPPPPIRGRGGRGGYGYPPDYGYEDYDYDYDYD 478
Qy 474 YHNYRGVYEDPYGYED-FQVGARGRGGRGARGA-APSRGRGAAPPGRAGYSQGGP-G 530
Db 479 YHNYRGVYEDPYGYEDYDGYAVRGRG-GRRGGRGAPPGRGRGAPPGRAGYSQGGP 537
Qy 531 SARGVRAGK-----RRRG 543
Db 538 PPRGSRGGRGGGPAQOQGRG 557

RESULT 2
PABP_SCHPO STANDARD; PRT; 653 AA.
AC P31209; P87135;
DT 01-JUL-1993 (Rel. 26, Created)
DT 15-JUL-1998 (Rel. 36, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE Polyadenylate-binding protein (Poly(A)-binding protein) (PABP) .
GN PAB1 OR PABP OR SPAC57A7.04C.
OS Schizosaccharomyces pombe (Fission yeast).
OC Eukaryota; Fungi; Ascomycota; Schizosaccharomycetes;
OC Schizosaccharomycetales; Schizosaccharomycetaceae;
OX NCBI_TaxID=4896;
RN [1]

```

```

RP SEQUENCE FROM N.A.
RC STRAIN=972;
RX MEDLINE=21848401; PubMed=11859360;
RA Wood V., Gwilliam R., Rajandream M.A., Lyne M., Lyne R., Stewart A.,
RA Sgouras J., Peat N., Hayles J., Baker S., Basham D., Bowman S.,
RA Brooks K., Brown D., Brown S., Chillingworth T., Churcher C.M.,
RA Collins M., Connor R., Cronin A., Davis P., Feltwell T., Fraser A.,
RA Gentles S., Goble A., Hamlin N., Harris D., Hidalgo J., Hodgson G.,
RA Holroyd S., Hornsby T., Howarth S., Huckle E.J., Hunt S., Jagels K.,
RA James K., Jones L., Jones M., Leather S., McDonald S., McLean J.,
RA Mooney P., Moule S., Mungall K., Murphy L., Niblett D., Odell C.,
RA Oliver K., O'Neill S., Pearson D., Quail M.A., Rabinowitsch E.,
RA Rutherford K., Rutter S., Saunders D., Seeger K., Sharp S.,
RA Skellon J., Simmonds M., Squares R., Squares S., Stevens K.,
RA Taylor K., Taylor R.G., Tivey A., Walsh S.V., Warren T., Whitehead S.,
RA Woodward J., Volkart G., Aert R., Robben J., Grymonprez B.,
RA Welfjens I., Vanstreels E., Rieger M., Schaefer M., Mueller-Auer S.,
RA Gabel C., Fuchs M., Fritz C., Holzer E., Moestl D., Hilbert H.,
RA Borzym K., Langer I., Beck A., Leirach H., Reinhardt R., Pohl T.M.,
RA Eger P., Zimmermann W., Wedler H., Wambutt R., Furnelle B.,
RA Goffeau A., Cadieu E., Dreano S., Gloux S., Lelaure V., Mottier S.,
RA Galibert F., Aves S.J., Xiang Z., Hunt C., Moore K., Hurst S.M.,
RA Lucas M., Rochet M., Gaillardin C., Tallada V.A., Garzon A., Thode G.,
RA Dominguez A., Cruzado L., Jimenez J., Sanchez M., del Rey F., Benito J.,
RA Daga R.R., Revuelta J.L., Moreno S., Armstrong J., Foreburg S.L.,
RA Cerrutti L., Lowe T., McCombie W.R., Paulsen I., Potashkin J.,
RA Shpakovski G.V., Ussery D., Barrell B.G., Nurse P.;
RT "The genome sequence of Schizosaccharomyces pombe.";
RL Nature 415:871-880(2002) .
RN [2]
RP SEQUENCE OF 15-628 FROM N.A.
RX MEDLINE=91260690; PubMed=1675426;
RA Burd C.G., Matunis E.L., Dreyfuss G.;
RT "The multiple RNA-binding domains of the mRNA poly(A)-binding protein
RT have different RNA-binding activities.";
RL Mol. Cell. Biol. 11:3419-3424(1991) .
CC -|- FUNCTION: BINDS THE POLY(A) TAIL OF MRNA.
CC -|- SUBCELLULAR LOCATION: Cytoplasmic.
CC -|- SIMILARITY: Contains 4 RNA recognition motif (RRM) domains.
CC -----
CC THIS SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; Z95396; CAB08762.1; -.
DR EMBL; M64603; AAA35320.1; -.
DR PIR; T38950; DNZPPA.
DR HSP; P11940; 1CVJ
DR GeneDB_Spombe; SPAC57A7.04c; -.
DR InterPro; IPR002004; PABP/HECT.
DR InterPro; IPR006515; PABP_1234.
DR InterPro; IPR000504; RNA_rec_mot.
DR Pfam; PF00658; PABP; 1.
DR Pfam; PF00076; Rrm; 4.
DR SMART; SM00517; PolyA; 1.
DR SMART; SM00360; RRM; 4.
DR TIGRFAMs; TIGR01628; PABP-1234; 1.
DR PROSITE; PS0102; RRM; 4.
DR PROSITE; PS0030; RRM_RNP_1; 3.
DR RNA-binding; Repeat.
DR DOMAIN 80 158 RNA-BINDING (RRM) 1.
DR DOMAIN 168 245 RNA-BINDING (RRM) 2.
DR DOMAIN 261 338 RNA-BINDING (RRM) 3.
DR DOMAIN 364 441 RNA-BINDING (RRM) 4.
DR CONFLICT 15 34 ESDVNTNNEAVESSTKEES -> MSLENSSTLSLCSNNT
DR CONFLICT 349 349 R -> A (IN REF. 2).
DR CONFLICT 518 526 TQFPAGGPA -> PLSSLLVR (IN REF. 2).
DR SEQUENCE 653 AA; 71512 MW; 7F8F5CAD69D0CFE1 CRC64;

```

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: January 28, 2004, 09:19:30 ; Search time 17 Seconds
(without alignments)

1507.621 Million cell updates/sec

Title: US-09-978-242-1

Perfect score: 2898

Sequence: 1 MATEHVNNGTEPMDTTSA.....RGGPSARGVGRACKRGGRS 545

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 127863 seqs, 47026705 residues

Total number of hits satisfying chosen parameters: 127863

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : SwissProt_41.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	2355	81.3	633	1 ROR HUMAN	O43390 homo sapien
2	287.5	9.9	653	1 PABP SCHPO	P13209 schizosacch
3	286	9.9	650	1 NUCLE XENLA	P20397 xenopus lae
4	274	9.5	345	1 SOD DROME	O08473 drosophila
5	260.5	9.0	644	1 PAB4 HUMAN	O13310 homo sapien
6	258.5	8.9	629	1 PAB2 ARATH	P42731 arabidopsis
7	258.5	8.9	713	1 NUCLE MESAU	P08199 mesocricetu
8	248	8.6	631	1 PAB3 SCHPO	Q9h361 homo sapien
9	247	8.5	500	1 GAR2 SCHPO	P41891 schizosacch
10	244	8.4	706	1 NUCLE HUMAN	P19338 homo sapien
11	242.5	8.4	636	1 PAB1 MOUSE	P29341 mus musculu
12	241.5	8.3	712	1 NUCLE RAT	P13383 rattus norv
13	240.5	8.3	636	1 PAB1 ARATH	P11940 homo sapien
14	236.5	8.2	668	1 PAB5 ARATH	Q05196 arabidopsis
15	236	8.1	576	1 PABP YEAST	P04147 saccharomyc
16	234.5	8.1	632	1 PABP DROME	P21187 drosophila
17	233	8.0	319	1 ROAL MOUSE	P49312 mus musculu
18	233	8.0	706	1 NUCLE MOUSE	P09405 mus musculu
19	232.5	8.0	414	1 NOP3 YEAST	Q01560 saccharomyc
20	232.5	8.0	633	1 PAB1 XENLA	P20965 xenopus lae
21	231	8.0	319	1 ROAL RAT	P04256 rattus norv
22	229	7.9	391	1 ROG HUMAN	P38159 homo sapien
23	229	7.9	694	1 NUCLE CHICK	P15771 gallus gall
24	227	7.8	522	1 PAB2 HUMAN	P15097 homo sapien
25	224.5	7.7	660	1 PAB3 ARATH	O64380 arabidopsis
26	223	7.7	319	1 ROAL MACMU	O28521 macaca mula
27	222.5	7.7	353	1 ROD RAT	O9j154 rattus norv
28	222	7.7	371	1 ROAL HUMAN	P09651 homo sapien
29	221.5	7.6	379	1 ROAL MOUSE	Q8b905 mus musculu
30	221	7.6	609	1 PABX ARATH	Q9zq48 arabidopsis
31	219.5	7.6	305	1 ROAO HUMAN	Q13151 homo sapien
32	219.5	7.6	355	1 ROD HUMAN	Q14103 homo sapien
33	218.5	7.5	424	1 S3B4_HUMAN	Q15427 homo sapien

34	218	7.5	385	1	RO32 XENLA	P51992 xenopus lae
35	215.5	7.4	307	1	SXL CHRRU	O97018 chrysomya r
36	214.5	7.4	388	1	ROG MOUSE	O35479 mus musculu
37	214	7.4	373	1	RO31 XENLA	P51968 xenopus lae
38	212.5	7.3	324	1	SXL MUSDO	O17310 musca domes
39	212	7.3	375	1	TIAR HUMAN	Q01085 homo sapien
40	210	7.2	346	1	RO21 XENLA	P51989 xenopus lae
41	209.5	7.2	382	1	PAB5 HUMAN	Q96du9 homo sapien
42	209	7.2	359	1	ELV2_HUMAN	Q12926 homo sapien
43	208	7.2	358	1	RO22_XENLA	P51990 xenopus lae
44	207.5	7.2	360	1	ELV2_MOUSE	Q60899 mus musculu
45	206.5	7.1	373	1	ELV4_RAT	O09032 rattus norv

ALIGNMENTS

RESULT 1

ID	ROR_HUMAN	STANDARD;	PRT;	633 AA.
AC	O43390;			
DT	16-OCT-2001 (Rel. 40, Created)			
DT	18-OCT-2001 (Rel. 40, Last sequence update)			
DT	28-FEB-2003 (Rel. 41, Last annotation update)			
DE	Heterogeneous nuclear ribonucleoprotein R (hnRNP R).			
GN	HNPR.			
OS	Homo sapiens (Human).			
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;			
OC	Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.			
OX	NCBI_TaxID=9606;			
RP	SEQUENCE FROM N.A.			
RN	MEDLINE=98083170; PubMed=9421497;			
RX	Hassfeld W., Chan E.K.L., Mathison D.A., Portman D., Dreyfuss G.,			
RA	Steiner G., Tan E.M.;			
RT	"Molecular definition of heterogeneous nuclear ribonucleoprotein R (hnRNP R) using autoimmune antibody: immunological relationship with hnRNP P.";			
RL	Nucleic Acids Res. 26:439-445(1998).			
CC	-1- FUNCTION: COMPONENT OF RIBONUCLEOSOMES, WHICH ARE COMPLEXES OF AT LEAST 20 OTHER DIFFERENT HETEROGENEOUS NUCLEAR RIBONUCLEOPROTEINS (HNPRNP). HNPRNP PLAY AN IMPORTANT ROLE IN PROCESSING OF PRECURSOR MRNA IN THE NUCLEUS.			
CC	-1- SUBCELLULAR LOCATION: Nuclear; nucleoplasm.			
CC	-1- SIMILARITY: Contains 3 RNA recognition motif (RRM) domains.			
CC	THIS SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See http://www.isb-sib.ch/announce/ or send an email to license@isb-sib.ch).			
CC	EMBL; AF000364; AAC39540.1; -			
DR	PIR; T02673; T02673.			
DR	HSSP; P09651; LHAI.			
DR	Genew; HGNC:5047; HNPRP.			
DR	GK; O43390; -			
DR	GO; GO:0005634; C:nucleus; TAS.			
DR	GO; GO:0008436; F:heterogeneous nuclear ribonucleoprotein; TAS.			
DR	GO; GO:0003723; F:RNA binding activity; TAS.			
DR	GO; GO:0006397; P:mRNA processing; TAS.			
DR	InterPro; IPR000535; hnRNP_R_Q.			
DR	InterPro; IPR000504; RNA_rec_mot.			
DR	Pfam; PF00076; rrm; 3.			
DR	SMART; SM00360; RRM; 3.			
DR	TIGRFAM; TIGR01648; hnRNP-R-Q; 1.			
DR	PROSITE; PS0102; RRM; 3.			
DR	PROSITE; PS00030; RRM RNP 1; 2.			
KW	Nuclear protein; RNA-binding; Ribonucleoprotein; Repeat.			
FT	DOMAIN 1 153 ASP/GLU-RICH (ACIDIC).			

Db 121 EAKIKALLERTGYTLDTVTGQKYGPPPSVYSGQPSVGTETIFVGKIPRDLFEDELVP 180
 Qy 181 LPEKAGPIWDLRLMMDPLTGLNRGYAFVFTCTKEAAQEAVKLYNNHIEIRSGKHIGVCISV 240
 Db 181 LPEKAGPIWDLRLMMDPLTGLNRGYAFVFTCTKEAAQEAVKLYNNHIEIRSGKHIGVCISV 240
 Qy 241 ANNRLFVGSIPKSKTKEQILEEFSKVTGLTDVILYHQDDKKKNGRCFCFLEYEDHKTAA 300
 Db 241 ANNRLFVGSIPKSKTKEQILEEFSKVTGLTDVILYHQDDKKKNGRCFCFLEYEDHKTAA 300
 Qy 301 QARRRLMSGKVKVNGVGTVEWADPIEDDPEVMAKVVLVFNRLANTVTEELEKAFSQ 360
 Db 301 QARRRLMSGKVKVNGVGTVEWADPIEDDPEVMAKVVLVFNRLANTVTEELEKAFSQ 360
 Qy 361 FGKLERVKLLKDYAFIHFDERDGAVKAMEEMNGKDLGENIEIVFAKPPDQKPKERKAOR 420
 Db 361 FGKLERVKLLKDYAFIHFDERDGAVKAMEEMNGKDLGENIEIVFAKPPDQKPKERKAOR 420
 Qy 421 QAAKNQMYDDYYGGPPHMPPTTRGRGRGGGGYPPDYGYEDYDYDYGYDYNHNRGG 480
 Db 421 QAAKNQMYDDYYGGPPHMPPTTRGRGRGGGGYPPDYGYEDYDYDYGYDYNHNRGG 480
 Qy 481 YEDPYGYEDFQVARGRGARGAAPSRRGGAAPPRGRAGYSQSGGPGSARGVRGARG 539
 Db 481 YEDPYGYEDFQVARGRGARGAAPSRRGGAAPPRGRAGYSQSGGPGSARGVRGARG 540
 Qy 540 -----RGRG 544
 Db 541 GAQQQRRGG 550

RESULT 2

Q9QYF4 ID Q9QYF4 PRELIMINARY; PRT; 561 AA.
 AC Q9QYF4
 DT 01-MAY-2000 (TrEMBLrel. 13, Created)
 DT 01-MAY-2000 (TrEMBLrel. 13, Last sequence update)
 DT 01-MAR-2003 (TrEMBLrel. 23, Last annotation update)
 DE SYNCRIP protein.
 GN NSAP1 OR SYNCRIP.
 OS Mus musculus (Mouse).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
 OX NCBI_TaxID=10090;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=DDY;
 RX MEDLINE=20200483; PubMed=10734137;
 RA Mizutani A., Fukuda M., Iwata K., Shiraiishi Y., Mikoshiba K.;
 RT "SYNCRIP, a cytoplasmic counterpart of heterogeneous nuclear
 ribonucleoprotein R, interacts with ubiquitous synaptotagmin
 RT isoforms.";
 RL J. Biol. Chem. 275:9823-9831 (2000).
 DR EMBL; AB035725; BAA88342.1; -.
 DR HSP; P19339; 2SXL.
 DR MGD; MGI:1891690; Neap1.
 DR InterPro; IPR006535; hRNP_R_Q.
 DR InterPro; IPR000504; RNA_rec_mot.
 DR Pfam; PF00076; rrm; 3.
 DR SMART; SM00360; RRM; 3.
 DR TIGRFAMs; TIGR01648; hRNP-R-Q; 1.
 DR PROSITE; PS0102; RRM; 3.
 DR PROSITE; PS00030; RRM_RNP_1; 2.
 SQ SEQUENCE 561 AA; 62544 MW; C0259C340146D16A CRC64;

Query Match 98.7%; Score 2860; DB 11; Length 561;
 Best Local Similarity 98.2%; Pred. No. 1.7e-191;
 Matches 539; Conservative 2; Mismatches 2; Indels 6; Gaps 1;
 Qy 1 MATEHVNGNGTEPMDDTTSVAVHSENFQTLDDAGLPQKVAEKLDEIYVAGLVVHSDLDLDER 60
 Db 1 MATEHVNGNGTEPMDDTTSVAVHSENFQTLDDAGLPQKVAEKLDEIYVAGLVVHSDLDLDER 60

Qy 61 AIEALKEFNEDGALAVLQFQKDSLSHVQNKSAFLCGVMKTYRQREKQCTKVADSSKGP 120
 Db 61 AIEALKEFNEDGALAVLQFQKDSLSHVQNKSAFLCGVMKTYRQREKQCTKVADSSKGP 120
 Qy 121 EAKIKALLERTGYTLDTVTGQKYGPPPSVYSGQPSVGTETIFVGKIPRDLFEDELVP 180
 Db 121 EAKIKALLERTGYTLDTVTGQKYGPPPSVYSGQPSVGTETIFVGKIPRDLFEDELVP 180
 Qy 181 LPEKAGPIWDLRLMMDPLTGLNRGYAFVFTCTKEAAQEAVKLYNNHIEIRSGKHIGVCISV 240
 Db 181 LPEKAGPIWDLRLMMDPLTGLNRGYAFVFTCTKEAAQEAVKLYNNHIEIRSGKHIGVCISV 240
 Qy 241 ANNRLFVGSIPKSKTKEQILEEFSKVTGLTDVILYHQDDKKKNGRCFCFLEYEDHKTAA 300
 Db 241 ANNRLFVGSIPKSKTKEQILEEFSKVTGLTDVILYHQDDKKKNGRCFCFLEYEDHKTAA 300
 Qy 301 QARRRLMSGKVKVNGVGTVEWADPIEDDPEVMAKVVLVFNRLANTVTEELEKAFSQ 360
 Db 301 QARRRLMSGKVKVNGVGTVEWADPIEDDPEVMAKVVLVFNRLANTVTEELEKAFSQ 360
 Qy 361 FGKLERVKLLKDYAFIHFDERDGAVKAMEEMNGKDLGENIEIVFAKPPDQKPKERKAOR 420
 Db 361 FGKLERVKLLKDYAFIHFDERDGAVKAMEEMNGKDLGENIEIVFAKPPDQKPKERKAOR 420
 Qy 421 QAAKNQMYDDYYGGPPHMPPTTRGRGRGGGGYPPDYGYEDYDYDYGYDYNHNRGG 480
 Db 421 QAAKNQMYDDYYGGPPHMPPTTRGRGRGGGGYPPDYGYEDYDYDYGYDYNHNRGG 480
 Qy 481 YEDPYGYEDFQVARGRGARGAAPSRRGGAAPPRGRAGYSQSGGPGSARGVRGARG 539
 Db 481 YEDPYGYEDFQVARGRGARGAAPSRRGGAAPPRGRAGYSQSGGPGSARGVRGARG 540
 Qy 540 -----RGRG 543
 Db 541 GAQQQRRGG 549

RESULT 3

Q9Y583 ID Q9Y583 PRELIMINARY; PRT; 562 AA.
 AC Q9Y583
 DT 01-NOV-1999 (TrEMBLrel. 12, Created)
 DT 01-NOV-1999 (TrEMBLrel. 12, Last sequence update)
 DT 01-MAR-2003 (TrEMBLrel. 23, Last annotation update)
 DE NSAP1 protein.
 GN NSAP1.
 OS Homo sapiens (Human).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
 OX NCBI_TaxID=9606;
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=99102562; PubMed=9847309;
 RA Harris C.E., Boden R.A., Astell C.R.;
 RT "A novel heterogeneous nuclear ribonucleoprotein-like protein
 RT interacts with NS1 of the minute virus of mice.";
 RL J. Virol. 73:72-80 (1999).
 DR EMBL; AF155568; AAD38198.1; -.
 DR HSP; P19339; 2SXL.
 DR InterPro; IPR006535; hRNP_R_Q.
 DR InterPro; IPR000504; RNA_rec_mot.
 DR Pfam; PF00076; rrm; 3.
 DR SMART; SM00360; RRM; 3.
 DR TIGRFAMs; TIGR01648; hRNP-R-Q; 1.
 DR PROSITE; PS0102; RRM; 3.
 DR PROSITE; PS00030; RRM_RNP_1; 2.
 SQ SEQUENCE 562 AA; 62656 MW; D48B582B00F946D CRC64;

Query Match 98.7%; Score 2860; DB 4; Length 562;
 Best Local Similarity 98.0%; Pred. No. 1.7e-191;
 Matches 539; Conservative 3; Mismatches 2; Indels 6; Gaps 1;
 Qy 1 MATEHVNGNGTEPMDDTTSVAVHSENFQTLDDAGLPQKVAEKLDEIYVAGLVVHSDLDLDER 60

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: January 28, 2004, 09:22:40 ; Search time 39 Seconds
(without alignments)
3606.120 Million cell updates/sec

Title: US-09-978-242-1
Perfect score: 2898
Sequence: 1 MATEHVNNGTPEPMDTSA.....RGPGSGARGVRACKRGGRS 545

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 830525 seqs, 258052604 residues

Total number of hits satisfying chosen parameters: 830525

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : SPTREMBL_23.*

- 1: sp_archaea.*
- 2: sp_bacteria.*
- 3: sp_fungi.*
- 4: sp_human.*
- 5: sp_invertebrate.*
- 6: sp_mammal.*
- 7: sp_mhc.*
- 8: sp_organelle.*
- 9: sp_phage.*
- 10: sp_plant.*
- 11: sp_rodent.*
- 12: sp_virus.*
- 13: sp_vertebrate.*
- 14: sp_unclassified.*
- 15: sp_rvirus.*
- 16: sp_bacteriaph.*
- 17: sp_cheap.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match %	Length	ID	Description
1	2861	98.7	562	11 Q8BGP1	Q8bgl mus musculus
2	2860	98.7	561	11 Q9QVF4	Q9qvf4 mus musculus
3	2860	98.7	562	4 Q9Y583	Q9y583 homo sapien
4	2860	98.7	627	11 Q8QGC2	Q8qgc2 mus musculus
5	2857	98.6	561	4 Q96LC1	Q96lc1 homo sapien
6	2857	98.6	623	4 Q60506	Q60506 homo sapien
7	2829	97.6	625	11 Q8B991	Q8b991 mus musculus
8	2647.5	91.4	527	4 Q8N599	Q8n599 homo sapien
9	2640.5	91.1	588	4 Q96LC2	Q96lc2 homo sapien
10	2533	87.4	558	11 Q91ZRO	Q91zr0 mus musculus
11	2487	85.8	491	11 Q8C5K6	Q8c5k6 mus musculus
12	2365.5	81.6	632	11 Q8VHM5	Q8vhm5 mus musculus
13	2343.5	80.9	636	4 Q9BV64	Q9bv64 homo sapien
14	2227	76.8	601	11 Q99KG1	Q99kg1 mus musculus
15	2023	69.8	410	4 Q81W78	Q81w78 homo sapien
16	1667	57.5	380	11 Q9CT37	Q9ct37 mus musculus

Q9vdi8 drosophila
Q8in47 drosophila
Q8in48 drosophila
Q95t44 drosophila
Q95t44 drosophila
Q9nld1 caenorhabdi
Q9xsr3 canis famli
Q8ni52 homo sapien
Q8bl32 mus musculu
Q8lwt8 mus musculu
Q8bze8 mus musculu
Q924k3 rattus norv
Q923k9 rattus norv
Q9nxd3 homo sapien
Q9nxd3 homo sapien
Q9nq94 homo sapien
Q9nq93 homo sapien
Q9nq93 homo sapien
Q8ch58 rattus norv
Q8ch57 rattus norv
Q8tby0 homo sapien
Q9nqx8 homo sapien
Q9asp6 arabidopsis
Q23093 arabidopsis
Q8w2r4 oryza sativ
Q9nxc9 homo sapien
Q9lxj8 arabidopsis
Q76494 dictyosteli
Q8rwq1 arabidopsis
Q8lyx4 homo sapien

ALIGNMENTS

RESULT 1

Q8BGP1 ID Q8BGP1 PRELIMINARY; PRT; 562 AA.

AC Q8BGP1;
DT 01-MAR-2003 (TREMREL. 23, Created)
DT 01-MAR-2003 (TREMREL. 23, Last sequence update)
DT 01-MAR-2003 (TREMREL. 23, Last annotation update)
DE NS1-associated protein 1.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=C57BL/6J; TISSUE=Body, and Embryo;
RX MEDLINE=22354683; PubMed=12466851;
RA The FANTOM Consortium,
RA the RIKEN Genome Exploration Research Group Phase I & II Team;
RT "Analysis of the mouse transcriptome based on functional annotation of
RL Nature 420:563-573(2002).
DR EMBL; AK034845; BAC28852.1; -
DR EMBL; AK077588; BAC36880.1; -
SQ SEQUENCE 562 AA; 62672 MW; 810BCFA286105377 CRC64;

Query Match 98.7%; Score 2861; DB 11; Length 562;
Best Local Similarity 98.0%; Pred. No. 1.5e-191;
Matches 539; Conservative 3; Mismatches 2; Indels 6; Gaps 1;
QY 1 MATEHVNNGTPEPMDTTSVHSENFQTLIDAGLPQKVAEKLDEIYVAGLVAHSOLDER 60
DB 1 MATEHVNNGTPEPMDTTSVHSENFQTLIDAGLPQKVAEKLDEIYVAGLVAHSOLDER 60
QY 61 ATEALKEFNEDGALAVLQFQKDSLSHVONKSAFLCGVMKTYRQKQTKVADSSKGPD 120
DB 61 ATEALKEFNEDGALAVLQFQKDSLSHVONKSAFLCGVMKTYRQKQTKVADSSKGPD 120
QY 121 EAKIKALLERTGYTLDTVTGQRKYGQPPDPSVYSGQPSVGTEIFVGKIPDLFEDELVP 180

source		1. .2079		/organisms="unknown"			
BASE COUNT		639 a		377 c		535 g	
ORIGIN				528 t			
Query Match		100.0%		Score 2079;		DB 6;	
Best Local Similarity		100.0%		Pred. No. 0;		Length 2079;	
Matches 2079;		Conservative 0;		Mismatches 0;		Indels 0;	
						Gaps 0;	
Qy	1	GGGCGCGCGCGCACCGGAGCGCGCTCGGAGCGGAGTGGAACTGGATCGGCTTTCGCTG	60				
Db	1	GGGCGCGCGCGCACCGGAGCGCGCTCGGAGCGGAGTGGAACTGGATCGGCTTTCGCTG	60				
Qy	61	CCAGCGCGTGAGCTTCGGCGCGCATTTTCAACAGCTCCACCTCGCGCCGGACACAGGGA	120				
Db	61	CCAGCGCGTGAGCTTCGGCGCGCATTTTCAACAGCTCCACCTCGCGCCGGACACAGGGA	120				
Qy	121	GCAGCGAGCAGCGGTTTCCCGCAACCCGATACCACTCGGACAGGATTTCTCGGCTCAGCC	180				
Db	121	GCAGCGAGCAGCGGTTTCCCGCAACCCGATACCACTCGGACAGGATTTCTCGGCTCAGCC	180				
Qy	181	CAACGGGGAGATCTCTGGAACATGGCTACAGACATGTTAATCGGAATGTAATGGA	240				
Db	181	CAACGGGGAGATCTCTGGAACATGGCTACAGACATGTTAATCGGAATGTAATGGA	240				
Qy	241	GCCCATGGATACTACTTCTGCAGTTATCCATTCAAGAAATTTTCAGACATTTGCTTGATGC	300				
Db	241	GCCCATGGATACTACTTCTGCAGTTATCCATTCAAGAAATTTTCAGACATTTGCTTGATGC	300				
Qy	301	TGGTTTACCACAGAAAGTTGCTGAAACCTAGATGAAATTTAAGTTCAGGGCTAGTTGC	360				
Db	301	TGGTTTACCACAGAAAGTTGCTGAAACCTAGATGAAATTTAAGTTCAGGGCTAGTTGC	360				
Qy	361	ACATAGTATTAGATGAAGAGCTATTGAGCTTTAAAGAAATTTCAATGAAGACGCTGC	420				
Db	361	ACATAGTATTAGATGAAGAGCTATTGAGCTTTAAAGAAATTTCAATGAAGACGCTGC	420				
Qy	421	ATTGGCAGTTCTTCAACAGTTTAAAGACAGTGATCTCTCATGCTTCAGAACAAAGTGC	480				
Db	421	ATTGGCAGTTCTTCAACAGTTTAAAGACAGTGATCTCTCTCATGCTTCAGAACAAAGTGC	480				
Qy	481	CTTTTATGTGGAGTCATGAAGACTTACAGCGAGAGAGAAAAAACAAGGACCAAGTAGC	540				
Db	481	CTTTTATGTGGAGTCATGAAGACTTACAGCGAGAGAGAAAAAACAAGGACCAAGTAGC	540				
Qy	541	AGATTCTAGTAAAGACCAAGATGAGGCAAAATTTAAGCACTCTTGGAAAGACAGGCTA	600				
Db	541	AGATTCTAGTAAAGACCAAGATGAGGCAAAATTTAAGCACTCTTGGAAAGACAGGCTA	600				
Qy	601	CACACTTGATGTGACCACTGGACAGAGAGAGTATGGAGGACCACTCCAGATTCGGTTTA	660				
Db	601	CACACTTGATGTGACCACTGGACAGAGAGAGTATGGAGGACCACTCCAGATTCGGTTTA	660				
Qy	661	TTCAGGTTCAGCAGCGCTTCTGTGTGGCACTGAGATATTGTGGGAAAGATCCCAAGATCT	720				
Db	661	TTCAGGTTCAGCAGCGCTTCTGTGTGGCACTGAGATATTGTGGGAAAGATCCCAAGATCT	720				
Qy	721	ATTTGAGGATGAACCTTGTTCCATTATTTGAGAAAGCTGGACCTATATGGGATCTTCGCT	780				
Db	721	ATTTGAGGATGAACCTTGTTCCATTATTTGAGAAAGCTGGACCTATATGGGATCTTCGCT	780				
Qy	781	AATGATGGATCCACTCACTGGTCTCAATAGAGGTTATGCGTTTGTCTACTTTTGTACAAA	840				
Db	781	AATGATGGATCCACTCACTGGTCTCAATAGAGGTTATGCGTTTGTCTACTTTTGTACAAA	840				
Qy	841	AGAAAGCAGCTCAGGAGGCTGTAAACCTGTATTAATCAATCAATGATTCGTTCTGAAACA	900				
Db	841	AGAAAGCAGCTCAGGAGGCTGTAAACCTGTATTAATCAATCAATGATTCGTTCTGAAACA	900				
Qy	901	TATTGGTGTCTGCACTCAGTTGCGCAACATAGGCTTTTGTGGGCTCTATTTCCTAAGAG	960				
Db	901	TATTGGTGTCTGCACTCAGTTGCGCAACATAGGCTTTTGTGGGCTCTATTTCCTAAGAG	960				
Qy	961	TAAACCAAGGAACAGATTTCTTGAAGATTTTAGCAAGTAAACAGAGGGTCTTACAGACGT	1020				
Db	961	TAAACCAAGGAACAGATTTCTTGAAGATTTTAGCAAGTAAACAGAGGGTCTTACAGACGT	1020				
Qy	1021	CATTTTATACCAACCAAGGATGACAAAGAAAAAAGAGGGCTTTTCTTCTTCAATA	1080				
Db	1021	CATTTTATACCAACCAAGGATGACAAAGAAAAAAGAGGGCTTTTCTTCTTCAATA	1080				
Qy	1081	TGAAGATCAAAAAACAGCTGCCAGCAAGGCGTAGTCTTAATGAGTGGTAAAGTCAAGGT	1140				
Db	1081	TGAAGATCAAAAAACAGCTGCCAGCAAGGCGTAGTCTTAATGAGTGGTAAAGTCAAGGT	1140				
Qy	1141	CTGGGGAATGTTGGAACTGTTGAATGGGCTGATCTCTATAGAAAGATCTCTGATCTGAGGT	1200				
Db	1141	CTGGGGAATGTTGGAACTGTTGAATGGGCTGATCTCTATAGAAAGATCTCTGATCTGAGGT	1200				
Qy	1201	TATGGCAAAAGGTAAAAGTGTCTGTACGCAACCTTGGCAATCTGTATACAGAAAGAT	1260				
Db	1201	TATGGCAAAAGGTAAAAGTGTCTGTACGCAACCTTGGCAATCTGTATACAGAAAGAT	1260				
Qy	1261	TTTAGAAAAAGGCAATTTAGTCAGTTTGGGAACTGGAAACGAGTGAAGATTTAAAGATTA	1320				
Db	1261	TTTAGAAAAAGGCAATTTAGTCAGTTTGGGAACTGGAAACGAGTGAAGATTTAAAGATTA	1320				
Qy	1321	TGCGTTCAATTCATTTTGTAGTGGAGAGATGCTGTCAAGGCTATGGAAAGATGAATGG	1380				
Db	1321	TGCGTTCAATTCATTTTGTAGTGGAGAGATGCTGTCAAGGCTATGGAAAGATGAATGG	1380				
Qy	1381	CAAGACTTGGAGGGAGAAATATTGAAATTTGTTTGGCAAGCCACCCAGATCAGAAAG	1440				
Db	1381	CAAGACTTGGAGGGAGAAATATTGAAATTTGTTTGGCAAGCCACCCAGATCAGAAAG	1440				
Qy	1441	GAAAGAAAAAGAAAGCTCAGAGGCAAGCAGCAAAAAATCAAATGTATGACCATTTACTACTA	1500				
Db	1441	GAAAGAAAAAGAAAGCTCAGAGGCAAGCAGCAAAAAATCAAATGTATGACCATTTACTACTA	1500				
Qy	1501	TTATGTTCCACTCATATATGCCCTCCAAACAGAGTTCGAGGCGTGGAGGTAGAGTGG	1560				
Db	1501	TTATGTTCCACTCATATATGCCCTCCAAACAGAGTTCGAGGCGTGGAGGTAGAGTGG	1560				
Qy	1561	TTATGATATCTCCAGATTTATGATATGAAGATTTATATGATTTATTTATGTTATGA	1620				
Db	1561	TTATGATATCTCCAGATTTATGATATGAAGATTTATATGATTTATTTATGTTATGA	1620				
Qy	1621	TTACCAATACTATCTGTGTGATATGAAGATCCATACTATGTTTATGAAGATTTTCAAGT	1680				
Db	1621	TTACCAATACTATCTGTGTGATATGAAGATCCATACTATGTTTATGAAGATTTTCAAGT	1680				
Qy	1681	TGAGGCTAGAGGAAGGGTGGTAGAGAGCAAGGGTGTCTCCATCCAGAGTCTGTGG	1740				
Db	1681	TGAGGCTAGAGGAAGGGTGGTAGAGAGCAAGGGTGTCTCCATCCAGAGTCTGTGG	1740				
Qy	1741	GGCTGTCTCTCCCGGTTAGACCGGTTATTCACAGAGAGGAGGCTCTCGATCAGCAAG	1800				
Db	1741	GGCTGTCTCTCCCGGTTAGACCGGTTATTCACAGAGAGGAGGCTCTCGATCAGCAAG	1800				
Qy	1801	AGCGTTTCGAGCAGGGAAGAGGGTTCGAGGCGGCTCTGACCTGTTTACAAATGAAGACTGA	1860				
Db	1801	AGCGTTTCGAGCAGGGAAGAGGGTTCGAGGCGGCTCTGACCTGTTTACAAATGAAGACTGA	1860				
Qy	1861	CTTGTCTATGTGGATTAACACAGAACTTCAGTGGAGTAAATGTTAAGGAAATCAAGCAA	1920				
Db	1861	CTTGTCTATGTGGATTAACACAGAACTTCAGTGGAGTAAATGTTAAGGAAATCAAGCAA	1920				
Qy	1921	CCTTAAATATGTCGGCTGATAGGACATATTTCTATGCAAGAGACCTCTCTATGAAGAT	1980				
Db	1921	CCTTAAATATGTCGGCTGATAGGACATATTTCTATGCAAGAGACCTCTCTATGAAGAT	1980				
Qy	1981	CATTGGAATCAAAATACGGGACATTTGAATCTTTGACCTTTGATATGAATTTCTTTAAC	2040				
Db	1981	CATTGGAATCAAAATACGGGACATTTGAATCTTTGACCTTTGATATGAATTTCTTTAAC	2040				
Qy	2041	AATTTTCTCTGCAGTCAAGTTTATTAACCTAAAAGCTACT	2079				

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: January 28, 2004, 22:58:47 ; Search time 5125 Seconds
(without alignments)
16595.339 Million cell updates/sec

Title: US-09-978-242-2

Perfect score: 2079

Sequence: 1 GGGCGCGCGCGCGCACCGGG.....GTTATTAACTAAAGCTACT 2079

Scoring table: IDENTITY_NUC

Gapop 10.0 , Gapext 1.0

Searched: 2888711 seqs, 2045481386 residues

Total number of hits satisfying chosen parameters: 5777422

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

GenEmbl:

- 1: gb_ba.*
- 2: gb_htg.*
- 3: gb_in.*
- 4: gb_om.*
- 5: gb_ov.*
- 6: gb_pat.*
- 7: gb_ph.*
- 8: gb_pl.*
- 9: gb_pr.*
- 10: gb_ro.*
- 11: gb_sts.*
- 12: gb_sy.*
- 13: gb_un.*
- 14: gb_vt.*
- 15: em_ba.*
- 16: em_fun.*
- 17: em_hum.*
- 18: em_in.*
- 19: em_mu.*
- 20: em_om.*
- 21: em_or.*
- 22: em_ov.*
- 23: em_pat.*
- 24: em_ph.*
- 25: em_pl.*
- 26: em_ro.*
- 27: em_sts.*
- 28: em_un.*
- 29: em_vi.*
- 30: em_htg_hum.*
- 31: em_htg_inv.*
- 32: em_htg_other.*
- 33: em_htg_mus.*
- 34: em_htg_pin.*
- 35: em_htg_rod.*
- 36: em_htg_mam.*
- 37: em_htg_vrt.*
- 38: em_sy.*
- 39: em_htgo_hum.*
- 40: em_htgo_mus.*
- 41: em_htgo_other.*

Pred. No. is the number of results predicted by chance to have a

score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB	ID	Description
1	2079	100.0	2079	6	AR177851	AR177851 Sequence
2	2012.6	96.8	2221	9	AF155568	AF155568 Homo sapi
3	1883.2	90.6	2208	10	BC041148	BC041148 Mus muscu
4	1848.4	88.9	103819	9	HSJ775C13	AL109618 Human DNA
5	1848.4	88.9	212659	2	AL390737	AL390737 Homo sapi
6	1826.2	87.8	3452	6	BD143848	BD143848 RNA-bindi
7	1826.2	87.8	3452	10	AB035725	AB035725 Mus muscu
8	1810.6	87.1	2932	9	AF037448	AF037448 Homo sapi
9	1809.2	87.1	3671	10	BC050079	BC050079 Mus muscu
10	1695.4	81.5	3319	9	BC032643	BC032643 Homo sapi
11	1619.8	77.9	1941	9	BC040844	BC040844 Homo sapi
12	1610.2	77.5	1686	9	AY034483	AY034483 Homo sapi
13	1610.2	77.5	1872	9	AY034481	AY034481 Homo sapi
14	1502.8	72.3	2373	10	AF093821	AF093821 Mus muscu
15	1463.4	70.4	3101	6	AX011753	AX011753 Sequence
16	1390.2	66.9	1767	9	AY034482	AY034482 Homo sapi
17	1329.2	63.9	2232	10	AF408434	AF408434 Mus muscu
18	1089.4	52.4	259969	2	AC118904	AC118904 Rattus no
19	1089.4	52.4	288143	2	AC109106	AC109106 Rattus no
20	935.2	45.0	2371	5	BC046902	BC046902 Danio rer
21	901.8	43.4	222602	2	AC122217	AC122217 Mus muscu
22	895	43.0	1899	10	AF441128	AF441128 Mus muscu
23	895	43.0	2725	10	BC038051	BC038051 Mus muscu
24	884.4	42.5	1899	10	AY184814	AY184814 Rattus no
25	882	42.4	2644	9	AF000364	AF000364 Homo sapi
26	864.6	41.6	2722	9	BC001449	BC001449 Homo sapi
27	852.8	41.0	2471	10	BC004679	BC004679 Mus muscu
28	846.4	40.7	235928	2	AC126148	AC126148 Rattus no
29	846.4	40.7	264178	2	AC098660	AC098660 Rattus no
30	792.4	38.1	934	6	AX013715	AX013715 Sequence
31	763	36.7	196488	2	AC084410	AC084410 Mus muscu
32	763	36.7	205520	2	AC079134	AC079134 Mus muscu
33	705.6	33.9	202374	10	AC122868	AC122868 Mus muscu
34	693.8	33.4	204962	2	AC127932	AC127932 Rattus no
35	693.8	33.4	225370	2	AC129693	AC129693 Rattus no
36	686.6	33.0	215801	2	AC140330	AC140330 Mus muscu
37	642.2	30.9	139996	9	AL161799	AL161799 Human DNA
38	482.8	23.2	2195	10	BC004001	BC004001 Mus muscu
39	445.2	21.4	1296	3	AK116214	AK116214 Ciona int
40	435.2	20.9	263129	2	AC098289	AC098289 Rattus no
41	424.6	20.4	121727	2	AC119124	AC119124 Rattus no
42	424.6	20.4	241150	2	AC105884	AC105884 Rattus no
43	407.2	19.6	265229	2	AC121468	AC121468 Rattus no
44	395.8	19.0	483	6	BD143847	BD143847 RNA-bindi
45	378.6	18.2	2606	3	AY058477	AY058477 Drosophil

ALIGNMENTS

RESULT 1	AR177851	2079 bp	DNA	linear	PAT 17-DEC-2001
LOCUS	AR177851	Sequence 2 from patent US 6313266.			
DEFINITION	AR177851				
ACCESSION	AR177851				
VERSION	AR177851.1	GI:17920206			
KEYWORDS					
SOURCE	Unknown.				
ORGANISM	Unknown.				
REFERENCE	1 (bases 1 to 2079)				
AUTHORS	Bandman,O., Yue,H., Corley,N.C. and Shah,P.				
TITLE	Human nucleolin-like protein				
JOURNAL	Patent: US 6313266-A 2 06-NOV-2001;				
FEATURES	Location/Qualifiers				

XX WPI: 1999-443600/37.
 DR P-PSDB; AAY24418.
 XX Nucleic acids encoding human nucleolin-like proteins useful for the
 PT diagnosis, prevention and treatment of disorders associated with
 PT abnormal nucleolin-like protein expression such as cancers,
 PT autoimmune disorders and Alzheimer's disease
 XX Claim 4; Fig 1; 33pp; English.
 XX The present sequence encodes human nucleolin-like peptide, designated
 CC HNL.P. HNL.P is the main protein component in the nucleolus of eukaryotic
 CC cells and is an essential part of ribosome biosynthesis and also plays
 CC an important role in importing proteins to the nucleus. HNL.P may be used
 CC in the diagnosis, prevention and treatment of disorders associated with
 CC abnormal expression of HNL.P. For example, it may be used to treat
 CC cancers (e.g. melanoma, breast cancer and prostate cancer), autoimmune
 CC disorders (e.g. autoimmune haemolytic anaemia and inflammatory bowel
 CC disease) and, in particular, Alzheimer's disease. It may also be used to
 CC study the function of the HNL.P peptides, the formation (biosynthesis) of
 CC ribosomes and the intake of proteins into the nucleus.
 XX SQ Sequence 2079 BP; 639 A; 377 C; 535 G; 528 T; 0 other;
 Query Match 100.0%; Score 2079; DB 20; Length 2079;
 Best Local Similarity 100.0%; Pred. No. 0;
 Matches 2079; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 Qy 1 GGGCGCGCGCGCACCGGAGCGCGCTCGGAGCGGAGTGGAACTGGATCGGTTTCTG 60
 Db 1 GGGCGCGCGCGCACCGGAGCGCGCTCGGAGCGGAGTGGAACTGGATCGGTTTCTG 60
 Qy 61 CCAGCGGCGTGGCTTCGGCGCGGATTTTCAACAGCTCCACTCGCGCGGACACAGGGA 120
 Db 61 CCAGCGGCGTGGCTTCGGCGCGGATTTTCAACAGCTCCACTCGCGCGGACACAGGGA 120
 Qy 121 GCAGCGAGCAGCGGTTTCCGCAACCGGATACCATCGGACAGGATTTCTCGCGCTCAGCC 180
 Db 121 GCAGCGAGCAGCGGTTTCCGCAACCGGATACCATCGGACAGGATTTCTCGCGCTCAGCC 180
 Qy 181 CAACGGGGAGATCTCTGAAACATGGCTACAGAACATGTTAATGGAATGGTACTGAAGA 240
 Db 181 CAACGGGGAGATCTCTGAAACATGGCTACAGAACATGTTAATGGAATGGTACTGAAGA 240
 Qy 241 GCCCATGGATACTACTTCTGAGTTATCCATTGAGAAATTTTCAGACATTTGCTTGATGC 300
 Db 241 GCCCATGGATACTACTTCTGAGTTATCCATTGAGAAATTTTCAGACATTTGCTTGATGC 300
 Qy 301 TGGTTTACACAGAAAGTTGCTGAAAACTAGATGAAATTTACGTTGCGAGGCTAGTTGC 360
 Db 301 TGGTTTACACAGAAAGTTGCTGAAAACTAGATGAAATTTACGTTGCGAGGCTAGTTGC 360
 Qy 361 ACATAGTGAATTTAGATGAAGAGCTATTGAAAGCTTTTAAAGAAATTTCAATGAAGCGTGC 420
 Db 361 ACATAGTGAATTTAGATGAAGAGCTATTGAAAGCTTTTAAAGAAATTTCAATGAAGCGTGC 420
 Qy 421 ATTGGCAGTTCTTCAACAGTTTAAAGACAGTATCTCTCATGTTTCAGAACAAAGTGC 480
 Db 421 ATTGGCAGTTCTTCAACAGTTTAAAGACAGTATCTCTCATGTTTCAGAACAAAGTGC 480
 Qy 481 CTTTATATGCGAGTCATGAAGCTTACAGCGAGAGAGAAACAAAGGACCAAGTAGC 540
 Db 481 CTTTATATGCGAGTCATGAAGCTTACAGCGAGAGAGAGAAACAAAGGACCAAGTAGC 540
 Qy 541 AGATTCTAGTAAAGGACAGATGAGGCAAAATTAAGGCACTCTTTGGAAGAACAGGCTA 600
 Db 541 AGATTCTAGTAAAGGACAGATGAGGCAAAATTAAGGCACTCTTTGGAAGAACAGGCTA 600
 Qy 601 CACACTTGATGTGACCACTGGACAGAGGAAGTATGGAGGACCACTTCAGATTCGGTTTA 660
 Db 601 CACACTTGATGTGACCACTGGACAGAGGAAGTATGGAGGACCACTTCAGATTCGGTTTA 660

Qy 661 TTCAGGTCAGCAGCCCTTCTGTTGGCACTGAGATATTGTTGGGAAAGATCCCAAGAGATCT 720
 Db 661 TTCAGGTCAGCAGCCCTTCTGTTGGCACTGAGATATTGTTGGGAAAGATCCCAAGAGATCT 720
 Qy 721 ATTGAGGATGAACCTTGTTCATTATTTGAGAAAGCTGGACCTATATGGATCTTCGTCT 780
 Db 721 ATTGAGGATGAACCTTGTTCATTATTTGAGAAAGCTGGACCTATATGGATCTTCGTCT 780
 Qy 781 AATGATGGATCCACTCACTCACTGCTCAATAGAGGTTATGCGTTTGTCACTTTTGTACAAA 840
 Db 781 AATGATGGATCCACTCACTCACTGCTCAATAGAGGTTATGCGTTTGTCACTTTTGTACAAA 840
 Qy 841 AGAAGCAGCTCAGAGGCTGTTAAACCTGATATAATCATGAATTCGTTCTGGAAACA 900
 Db 841 AGAAGCAGCTCAGAGGCTGTTAAACCTGATATAATCATGAATTCGTTCTGGAAACA 900
 Qy 901 TATTGGTGTGCTGATCTCAGTTCGCCCAACATAGGCTTTTGTGGGCTCTATTCCTAAGAG 960
 Db 901 TATTGGTGTGCTGATCTCAGTTCGCCCAACATAGGCTTTTGTGGGCTCTATTCCTAAGAG 960
 Qy 961 TAAACCAAGGAACAGATCTTCAAGAAATTTAGCAAGTAAACAGAGGCTCTTACAGACGT 1020
 Db 961 TAAACCAAGGAACAGATCTTCAAGAAATTTAGCAAGTAAACAGAGGCTCTTACAGACGT 1020
 Qy 1021 CATTTTATACCAACCGGATGACAAAGAAAAAACAAGAGGCTTTTGTCTTCTGAATA 1080
 Db 1021 CATTTTATACCAACCGGATGACAAAGAAAAAACAAGAGGCTTTTGTCTTCTGAATA 1080
 Qy 1081 TGAAGATCAAAACAGCTGCCCGAGGCGTAGTTAATGAGTGGTAAAGTCAAGGT 1140
 Db 1081 TGAAGATCAAAACAGCTGCCCGAGGCGTAGTTAATGAGTGGTAAAGTCAAGGT 1140
 Qy 1141 CTGGGGGAATGTTGGAACCTGTTCAATGGGCTGATCTTATAGAAGATCTCTGATCCTGAGGT 1200
 Db 1141 CTGGGGGAATGTTGGAACCTGTTCAATGGGCTGATCTTATAGAAGATCTCTGATCCTGAGGT 1200
 Qy 1201 TATGGCAAGGTAAGTGTCTGTTGTACCAACCTTGGCAATATCTGTATAACAGAGAT 1260
 Db 1201 TATGGCAAGGTAAGTGTCTGTTGTACCAACCTTGGCAATATCTGTATAACAGAGAT 1260
 Qy 1261 TTTAGAAAGGCAATTTAGTCAGTTTGGGAACTGGAACTGGAACTGGAACTGGAACTGGAA 1320
 Db 1261 TTTAGAAAGGCAATTTAGTCAGTTTGGGAACTGGAACTGGAACTGGAACTGGAACTGGAA 1320
 Qy 1321 TGGTTTCACTTCTTATGATGAGCGAGATGCTGTCTCAAGCTTATGGAAGAAATGAATGG 1380
 Db 1321 TGGTTTCACTTCTTATGATGAGCGAGATGCTGTCTCAAGCTTATGGAAGAAATGAATGG 1380
 Qy 1381 CAAAGACTTGGAGGAGAAATATTGAAATTTGTTTGGCAAGCCACAGATCAGAAAAG 1440
 Db 1381 CAAAGACTTGGAGGAGAAATATTGAAATTTGTTTGGCAAGCCACAGATCAGAAAAG 1440
 Qy 1441 GAAAGAAAGAAAGCTCAGAGGCAAGCAGCAAAATATGATGACGATTTACTACTA 1500
 Db 1441 GAAAGAAAGAAAGCTCAGAGGCAAGCAGCAAAATATGATGACGATTTACTACTA 1500
 Qy 1501 TTATGCTCCACCTCATATGCTCCCAACAGAGGTCGAGGCGCTGAGGAGTGAAGGTGG 1560
 Db 1501 TTATGCTCCACCTCATATGCTCCCAACAGAGGTCGAGGCGCTGAGGAGTGAAGGTGG 1560
 Qy 1561 TTATGATATCTCCAGATTTATTTGATATGAAGATTTATTTATGATTTATGATTTATGA 1620
 Db 1561 TTATGATATCTCCAGATTTATTTGATATGAAGATTTATTTATGATTTATGATTTATGA 1620
 Qy 1621 TTACCATTAATCTGCTGCTGATATGAAGATCCATCTATGTTTATGAAGATTTTCAAGT 1680
 Db 1621 TTACCATTAATCTGCTGCTGATATGAAGATCCATCTATGTTTATGAAGATTTTCAAGT 1680
 Qy 1681 TGGAGCTAGAGGAGGCGTGTAGAGGAGCAAGGGTGTCTCTCATCCAGAGGTCGTTGG 1740
 Db 1681 TGGAGCTAGAGGAGGCGTGTAGAGGAGCAAGGGTGTCTCTCATCCAGAGGTCGTTGG 1740
 Qy 1741 GGCTGCTCTCTCCCGCGGTAGAGCGCGTTTATTTCAAGAGAGGAGGTCTCTGATCAGCAAG 1800

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: January 28, 2004, 22:57:52 ; Search time 411 Seconds
(without alignments)
13654.822 Million cell updates/sec

Title: US-09-978-242-2

Perfect score: 2079

Sequence: 1 GGGCGCGCGCGCGACCGG.....GTTATTAACTAAAGCTACT 2079

Scoring table: IDENTITY NUC

Gapop 10.0 , Gapext 1.0

Searched: 2552756 seqs, 1349719017 residues

Total number of hits satisfying chosen parameters: 5105512

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Listing first 45 summaries

Database :

N_Geneseq_19Jun03:*

- 1: /SIDS1/gcgdata/geneseq/geneseq-emb1/NA1980.DAT:*
- 2: /SIDS1/gcgdata/geneseq/geneseq-emb1/NA1981.DAT:*
- 3: /SIDS1/gcgdata/geneseq/geneseq-emb1/NA1982.DAT:*
- 4: /SIDS1/gcgdata/geneseq/geneseq-emb1/NA1983.DAT:*
- 5: /SIDS1/gcgdata/geneseq/geneseq-emb1/NA1984.DAT:*
- 6: /SIDS1/gcgdata/geneseq/geneseq-emb1/NA1985.DAT:*
- 7: /SIDS1/gcgdata/geneseq/geneseq-emb1/NA1986.DAT:*
- 8: /SIDS1/gcgdata/geneseq/geneseq-emb1/NA1987.DAT:*
- 9: /SIDS1/gcgdata/geneseq/geneseq-emb1/NA1988.DAT:*
- 10: /SIDS1/gcgdata/geneseq/geneseq-emb1/NA1989.DAT:*
- 11: /SIDS1/gcgdata/geneseq/geneseq-emb1/NA1990.DAT:*
- 12: /SIDS1/gcgdata/geneseq/geneseq-emb1/NA1991.DAT:*
- 13: /SIDS1/gcgdata/geneseq/geneseq-emb1/NA1992.DAT:*
- 14: /SIDS1/gcgdata/geneseq/geneseq-emb1/NA1993.DAT:*
- 15: /SIDS1/gcgdata/geneseq/geneseq-emb1/NA1994.DAT:*
- 16: /SIDS1/gcgdata/geneseq/geneseq-emb1/NA1995.DAT:*
- 17: /SIDS1/gcgdata/geneseq/geneseq-emb1/NA1996.DAT:*
- 18: /SIDS1/gcgdata/geneseq/geneseq-emb1/NA1997.DAT:*
- 19: /SIDS1/gcgdata/geneseq/geneseq-emb1/NA1998.DAT:*
- 20: /SIDS1/gcgdata/geneseq/geneseq-emb1/NA1999.DAT:*
- 21: /SIDS1/gcgdata/geneseq/geneseq-emb1/NA2000.DAT:*
- 22: /SIDS1/gcgdata/geneseq/geneseq-emb1/NA2001A.DAT:*
- 23: /SIDS1/gcgdata/geneseq/geneseq-emb1/NA2001B.DAT:*
- 24: /SIDS1/gcgdata/geneseq/geneseq-emb1/NA2002.DAT:*
- 25: /SIDS1/gcgdata/geneseq/geneseq-emb1/NA2003.DAT:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	2079	100.0	2079	20	Human nucleolin-li
2	2079	100.0	2079	20	Human cDNA encodin
3	1844	88.7	4231	21	Human prostate can
4	1826.2	87.8	3452	24	Nucleotide sequenc
5	1777.6	85.5	2680	21	Human RNA-associat
6	1731	83.3	2407	23	DNA encoding novel
7	1463.4	70.4	3101	20	Human prostate tum
8	882	42.4	2673	21	Human secreted pro

	9	882	42.4	3159	25	ACA03964	CDNA downregulated
c	10	871.6	41.9	2671	24	ABL35808	CDNA sequence #199
	11	839.8	40.4	2714	22	ABL26717	Human breast cancer
	12	833.6	40.1	2700	22	AH33268	Human colon cancer
	13	792.4	38.1	934	20	AA242164	Human normal bladd
	14	730	35.1	1024	21	AA98058	Human colon cancer
	15	498.8	24.0	612	21	AA280570	Human colon cancer
	16	487	23.4	1907	24	AAK35125	Human cDNA encodin
	17	482.8	23.2	560	24	ABQ58093	Human colon cancer
	18	425	20.4	464	22	ABA51954	Human foetal liver
	19	425	20.4	464	22	ABA21772	Probe #238 for gen
	20	425	20.4	464	22	AAK00240	Human brain expres
	21	425	20.4	464	22	AAK25682	Human bone marrow
	22	425	20.4	464	22	AAI10309	Probe #242 for gen
	23	425	20.4	464	22	AAI31562	Probe #248 used to
	24	425	20.4	464	22	AAI00245	Probe #236 used to
	25	425	20.4	464	23	ABS25267	Human liver single
	26	425	20.4	464	24	ABS00260	Human genome-deriv
	27	395.8	19.0	483	24	ABL59087	Nucleotide sequenc
	28	371.8	17.9	384	19	AAV54582	Human secretory pr
	29	371.8	17.9	384	20	AA225601	Human secreted pro
	30	354.6	17.1	747	24	ABS77417	Frog embryonic gen
	31	287.8	13.8	591	22	AAH35088	Human colon cancer
	32	275.2	13.2	700	24	ABS77066	Frog embryonic gen
	33	272.8	13.1	382	22	ABA46733	Human breast cell
	34	272.8	13.1	382	22	ABA64613	Human foetal liver
	35	272.8	13.1	382	22	ABA31740	Probe #10206 for g
	36	272.8	13.1	382	22	AAK13051	Human brain expres
	37	272.8	13.1	382	22	AAK38781	Human bone marrow
	38	272.8	13.1	382	22	AAI19587	Probe #9520 for ge
	39	272.8	13.1	382	22	AAI44783	Probe #13469 used
	40	272.8	13.1	382	22	AAI05307	Probe #5298 used t
	41	272.8	13.1	382	23	ABS38364	Human liver single
	42	272.8	13.1	382	24	ABS12860	Human genome-deriv
	43	262.2	12.6	342	14	AAQ59054	Human brain expres
	44	261.6	12.6	342	14	AAQ39642	Expressed Sequence
	45	261	12.6	307	20	AAV86918	EST clone BK260.

ALIGNMENTS

RESULT 1	AA90183	standard; cDNA; 2079 BP.
ID	AA90183	
XX	AA90183	
AC	AA90183	
DT	21-SEP-1999	(first entry)
XX	Human nucleolin-like protein encoding cDNA.	
DE	Human, nucleolin-like protein; HNLIP; cancer; melanoma; breast cancer;	
KW	prostate cancer; autoimmune disorder; autoimmune haemolytic anaemia;	
KW	inflammatory bowel disease; Alzheimer's disease; ss.	
XX	Homo sapiens.	
OS	Homo sapiens.	
XX	Key	Location/Qualifiers
FH	CDS	203..1840
FT		/*tag= a
XX		
PN	US932475-A.	
XX		
PD	03-AUG-1999.	
XX		
PF	12-DEC-1997;	97US-0990114.
XX		
PR	12-DEC-1997;	97US-0990114.
XX		
PA	(INCY-) INCYTE PHARM INC.	
XX		
PI	Bandman O, Corley NC, Shah P, Yue H;	

Best Local Similarity 100.0%; Pred. No. 0; Matches 2079; Conservative 0; Mismatches 0; Indels 0; Gaps 0;									
Qy	1	GGGCGCGCGCGCACCGGAGCGCGCTCGAGGCGAGTGGAACTGGATCGGTTTCTG	60						
Db	1	GGGCGCGCGCGCACCGGAGCGCGCTCGAGGCGAGTGGAACTGGATCGGTTTCTG	60						
Qy	61	CCAGCGCGTGAGCTTCGGCGCGCATTTTCAACAGCTCCACTCGCGCGCGACACAGGGA	120						
Db	61	CCAGCGCGTGAGCTTCGGCGCGCATTTTCAACAGCTCCACTCGCGCGCGACACAGGGA	120						
Qy	121	GCAGCGAGCAGCGTTTCCCGCAACCCGATACCATCGGACAGGATTTCTCGGCTCAGCC	180						
Db	121	GCAGCGAGCAGCGTTTCCCGCAACCCGATACCATCGGACAGGATTTCTCGGCTCAGCC	180						
Qy	181	CAACGGGAGATCTCTGGAAACATGGCTACAGACATGTTAATGGAAATGGTACTGAAGA	240						
Db	181	CAACGGGAGATCTCTGGAAACATGGCTACAGACATGTTAATGGAAATGGTACTGAAGA	240						
Qy	241	GCCCATGGATACTTCTCGAGTTATCCATTCAAGAAATTTTCAGACATTTGCTTGATGC	300						
Db	241	GCCCATGGATACTTCTCGAGTTATCCATTCAAGAAATTTTCAGACATTTGCTTGATGC	300						
Qy	301	TGGTTTACCAGAAAGTTGCTGAAAACTAGATGAATTTTACGTTGCGGCTAGTTGC	360						
Db	301	TGGTTTACCAGAAAGTTGCTGAAAACTAGATGAATTTTACGTTGCGGCTAGTTGC	360						
Qy	361	ACATAGTGATTTAGATGAAGAGCTTATGAAGCTTTAAAGAAATTTCAATGAAGACGGTGC	420						
Db	361	ACATAGTGATTTAGATGAAGAGCTTATGAAGCTTTAAAGAAATTTCAATGAAGACGGTGC	420						
Qy	421	ATTGGCAGTCTTCAACAGTTTAAAGACAGTGATCTCTCTCATGTTTCAGAACAAAGTGC	480						
Db	421	ATTGGCAGTCTTCAACAGTTTAAAGACAGTGATCTCTCTCATGTTTCAGAACAAAGTGC	480						
Qy	481	CTTTTATGTGGAGTCATGAAGCTTACAGCAGAGAGAAAAACAAGGACCAAGTAGC	540						
Db	481	CTTTTATGTGGAGTCATGAAGCTTACAGCAGAGAGAAAAACAAGGACCAAGTAGC	540						
Qy	541	AGATTCTAGTAAAGGACAGATGAGGCAAAATTAAGGCATCTTGGAAAGACAGGCTA	600						
Db	541	AGATTCTAGTAAAGGACAGATGAGGCAAAATTAAGGCATCTTGGAAAGACAGGCTA	600						
Qy	601	CACACTGTGACCTGACAGAGAGAGTATGGAGGACCACTCCAGATTCCGTTTA	660						
Db	601	CACACTGTGACCTGACAGAGAGAGTATGGAGGACCACTCCAGATTCCGTTTA	660						
Qy	661	TTCAGGTCAGCAGCCTTCTGTTGGCAGTGAATATTTTGGGAAAGATCCCAAGAGATCT	720						
Db	661	TTCAGGTCAGCAGCCTTCTGTTGGCAGTGAATATTTTGGGAAAGATCCCAAGAGATCT	720						
Qy	721	ATTTGAGGATGAATTTGTTTCAATTTTGAAGAGCTGGACCTATATGGGATCTTCGTC	780						
Db	721	ATTTGAGGATGAATTTGTTTCAATTTTGAAGAGCTGGACCTATATGGGATCTTCGTC	780						
Qy	781	AATGATGATCCACTCAGTCTCAATAGAGTTATCGGTTGTCACCTTTTGTACAAA	840						
Db	781	AATGATGATCCACTCAGTCTCAATAGAGTTATCGGTTGTCACCTTTTGTACAAA	840						
Qy	841	AGAAGCAGCTCAGGAGGCTGTTTAACTGTATAATCATGAAATTCGTTCTGGAAAAACA	900						
Db	841	AGAAGCAGCTCAGGAGGCTGTTTAACTGTATAATCATGAAATTCGTTCTGGAAAAACA	900						
Qy	901	TATTGGTCTGTCATCTCAGTTGCCAACAATAGGCTTTTGTGGGCTCTATTCCTAAGAG	960						
Db	901	TATTGGTCTGTCATCTCAGTTGCCAACAATAGGCTTTTGTGGGCTCTATTCCTAAGAG	960						
Qy	961	TAAACCAAGGAACAGATTTCTTGAAGATTTTAGCAAGTAAACAGAGGCTTTACAGAGCT	1020						
Db	961	TAAACCAAGGAACAGATTTCTTGAAGATTTTAGCAAGTAAACAGAGGCTTTACAGAGCT	1020						
Qy	1021	CATTTTATACCACCGGATGACAGAAAAAACAAGAGGCTTTTGTCTTTCTTGAATA	1080						

Db	1021	CATTTTATACCACCGGATGACAGAAAAAACAAGAGGCTTTTGTCTTTCTTGAATA	1080						
Qy	1081	TGAAGATACAAAACAGCTGCCAGGCAAGGCGTAGGTTAACTAGTGGTAAAGTCAAGGT	1140						
Db	1081	TGAAGATACAAAACAGCTGCCAGGCAAGGCGTAGGTTAACTAGTGGTAAAGTCAAGGT	1140						
Qy	1141	CTGGGGAAATGTTGGAACTGTTGAAATGGGCTGATCTATAGAAGATCTCTGATCCTGAGGT	1200						
Db	1141	CTGGGGAAATGTTGGAACTGTTGAAATGGGCTGATCTATAGAAGATCTCTGATCCTGAGGT	1200						
Qy	1201	TATGGCAAGGTTAAAGTGTCTGTACGCAACCTTGGCAATCTGTAACAGAGAGAT	1260						
Db	1201	TATGGCAAGGTTAAAGTGTCTGTGTACGCAACCTTGGCAATCTGTAACAGAGAGAT	1260						
Qy	1261	TTTATAGAAAGGCAATTTAGTCAGTTTCGGGAACTGGAAACGAGTGAAGAAATTA	1320						
Db	1261	TTTATAGAAAGGCAATTTAGTCAGTTTCGGGAACTGGAAACGAGTGAAGAAATTA	1320						
Qy	1321	TGCGTTCAATTCATTTTGTATGAGCGAGATGGTGTCTCAAGGCTTATGGAGAAATGAATGG	1380						
Db	1321	TGCGTTCAATTCATTTTGTATGAGCGAGATGGTGTCTCAAGGCTTATGGAGAAATGAATGG	1380						
Qy	1381	CAAGACTTGGAGGAGAAATATTGAAATTTGTTTTCGCAAGCCACCCAGATCAGAAAAG	1440						
Db	1381	CAAGACTTGGAGGAGAAATATTGAAATTTGTTTTCGCAAGCCACCCAGATCAGAAAAG	1440						
Qy	1441	GAAAGAAAGAAAGCTCAGAGGCAAGCAGCAAAATTTGTTTTCGCAAGCCACCCAGATCAGAAAAG	1500						
Db	1441	GAAAGAAAGAAAGCTCAGAGGCAAGCAGCAAAATTTGTTTTCGCAAGCCACCCAGATCAGAAAAG	1500						
Qy	1501	TTATGGTCCACTCATATATGAGGCTTCAACAGAGGTCGAGGCGTGGAGGTAGAGGTGG	1560						
Db	1501	TTATGGTCCACTCATATATGAGGCTTCAACAGAGGTCGAGGCGTGGAGGTAGAGGTGG	1560						
Qy	1561	TTATGATATCTCCAGATTTATATGATATGAAGATTTATGATTTATGTTATGA	1620						
Db	1561	TTATGATATCTCCAGATTTATATGATATGAAGATTTATGATTTATGTTATGA	1620						
Qy	1621	TTACCAATACTATCGTGGTGATATGAAGATCCATCTACTATGTTTATGAAGATTTTCAAGT	1680						
Db	1621	TTACCAATACTATCGTGGTGATATGAAGATCCATCTACTATGTTTATGAAGATTTTCAAGT	1680						
Qy	1681	TGGAGCTTAGAGGAGGCGTGTAGAGGAGCAAGGCGTGTCTTCCATCCAGAGGTCTGG	1740						
Db	1681	TGGAGCTTAGAGGAGGCGTGTAGAGGAGCAAGGCGTGTCTTCCATCCAGAGGTCTGG	1740						
Qy	1741	GGCTGTCTCTCCCGGCTAGAGCCGTTTATTCACAGAGGAGGCTCCTCGATCAGCAAG	1800						
Db	1741	GGCTGTCTCTCCCGGCTAGAGCCGTTTATTCACAGAGGAGGCTCCTCGATCAGCAAG	1800						
Qy	1801	AGGCGTTTCAGAGGAGGAAAGGCGTCCAGGCGGCTCTGACCTGTTTACAAATGAAGACTGA	1860						
Db	1801	AGGCGTTTCAGAGGAGGAAAGGCGTCCAGGCGGCTCTGACCTGTTTACAAATGAAGACTGA	1860						
Qy	1861	CTTGCTATGTGGGATTTACACAGAGCTTCAGTGGAGTAAATGGTAAAGAAATCAAGCAA	1920						
Db	1861	CTTGCTATGTGGGATTTACACAGAGCTTCAGTGGAGTAAATGGTAAAGAAATCAAGCAA	1920						
Qy	1921	CCTTAAATATGTCGGCTGTATAGGAGCATTTCTATTCGAGAGACCTTCTATGAAGAT	1980						
Db	1921	CCTTAAATATGTCGGCTGTATAGGAGCATTTCTATTCGAGAGACCTTCTATGAAGAT	1980						
Qy	1981	CATGGAATCAAAATACGGGACATTTGAACATACTTTGACCTTTTGAATGAATTTCTTAAAC	2040						
Db	1981	CATGGAATCAAAATACGGGACATTTGAACATACTTTGACCTTTTGAATGAATTTCTTAAAC	2040						
Qy	2041	AATTTTCTCTGCAAGTCAAGTATTAACTTAAAGCTACT 2079							
Db	2041	AATTTTCTCTGCAAGTCAAGTATTAACTTAAAGCTACT 2079							

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: January 29, 2004, 00:15:58 ; Search time 109 Seconds
(without alignments)
8418.675 Million cell updates/sec

Title: US-09-978-242-2
Perfect score: 2079
Sequence: 1 GGGCGCGCGCGCGCACCGG.....GTTATTAACTAAAGCTACT 2079

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 569978 seqs, 220691566 residues

Total number of hits satisfying chosen parameters: 1139956

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Issued Patents NA: *
1: /cgn2_6/ptodata/2/ina/5A COMB.seq: *
2: /cgn2_6/ptodata/2/ina/5B COMB.seq: *
3: /cgn2_6/ptodata/2/ina/6A COMB.seq: *
4: /cgn2_6/ptodata/2/ina/6B COMB.seq: *
5: /cgn2_6/ptodata/2/ina/PTUS COMB.seq: *
6: /cgn2_6/ptodata/2/ina/backfiles1.seq: *

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES												
Result No.	Score	Query Match	Length	ID	Description							
1	2079	100.0	2079	2	US-08-990-114-2	Sequence 2, Appli						
2	2079	100.0	2079	4	US-09-241-333-2	Sequence 2, Appli						
3	498.8	24.0	612	3	US-09-328-111-654	Sequence 654, App						
4	371.8	17.9	384	2	US-09-014-969-3	Sequence 3, Appli						
5	65.6	3.2	7218	1	US-08-232-463-14	Sequence 14, Appli						
6	48.4	2.3	2361	4	US-09-152-060-28	Sequence 28, Appli						
7	47.8	2.3	18596	3	US-09-318-448-11	Sequence 11, Appli						
8	47.8	2.3	18597	4	US-09-962-665-8	Sequence 8, Appli						
9	47.2	2.3	2277	1	US-08-676-967-2	Sequence 2, Appli						
10	47.2	2.3	2277	1	US-08-676-974-2	Sequence 2, Appli						
11	47.2	2.3	2277	2	US-09-098-487-2	Sequence 2, Appli						
12	46.4	2.2	1894	2	US-08-935-450-7	Sequence 7, Appli						
13	44.8	2.2	309	1	US-08-209-747-3	Sequence 3, Appli						
14	44.8	2.2	309	1	US-08-458-298-3	Sequence 3, Appli						
15	44.2	2.1	1505	1	US-07-915-246-1	Sequence 1, Appli						
16	44.2	2.1	4673	1	US-07-638-431-1	Sequence 1, Appli						
17	44.2	2.1	4673	5	PCT-US92-00018-1	Sequence 1, Appli						
18	43.6	2.1	10640	4	US-09-417-4850-5	Sequence 5, Appli						
19	43.2	2.1	2042	2	US-08-911-434A-2	Sequence 2, Appli						
20	42.6	2.0	527	3	US-09-347-833-5	Sequence 5, Appli						
21	42.4	2.0	5340	4	US-09-627-122-21	Sequence 21, Appli						
22	42.4	2.0	8920	2	US-08-446-855A-1	Sequence 1, Appli						
23	42.4	2.0	8920	3	US-09-150-741-1	Sequence 1, Appli						
24	42.2	2.0	2793	1	US-08-209-747-1	Sequence 1, Appli						
25	42.2	2.0	2793	1	US-08-458-298-1	Sequence 1, Appli						
26	42.2	2.0	1664976	4	US-08-916-421B-1	Sequence 1, Appli						
27	41.8	2.0	5340	4	US-09-627-122-21	Sequence 21, Appli						

Sequence 4, Appli
Sequence 4, Appli
Sequence 3, Appli
Sequence 3, Appli
Sequence 3, Appli
Sequence 1, Appli
Sequence 1, Appli
Sequence 20, Appli
Sequence 20, Appli
Sequence 7, Appli
Sequence 19, Appli
Sequence 3, Appli
Sequence 2161, Ap
Sequence 1, Appli
Sequence 2, Appli
Sequence 4, Appli

ALIGNMENTS

RESULT 1
US-08-990-114-2
; Sequence 2, Application US/08990114
; Patent No. 5932475
; GENERAL INFORMATION:
; APPLICANT: Bandman, Olga
; APPLICANT: Yue, Henry
; APPLICANT: Corley, Neil C.
; APPLICANT: Shah, Purvi
; TITLE OF INVENTION: HUMAN NUCLEOLIN-LIKE PROTEIN
; NUMBER OF SEQUENCES: 3
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Incyte Pharmaceuticals, Inc.
; STREET: 3174 Porter Drive
; CITY: Palo Alto
; STATE: CA
; COUNTRY: USA
; ZIP: 94304
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette
; OPERATING SYSTEM: DOS
; SOFTWARE: FastSeq for Windows Version 2.0
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/990,114
; FILING DATE: Herewith
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER:
; FILING DATE:
; ATTORNEY/AGENT INFORMATION:
; NAME: Billings, Lucy J.
; REGISTRATION NUMBER: 36,749
; REFERENCE/DOCKET NUMBER: PF-0451 US
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 650-855-0555
; TELEFAX: 650-845-4166
; TELEX:
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 2079 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; IMMEDIATE SOURCE:
; LIBRARY: TLYMNOT05
; CLONE: 2809795
; US-08-990-114-2

Query Match 100.0%; Score 2079; DB 2; Length 2079;

```
;
; TELEFAX: 650-845-4166
; TELEX: <Unknown>
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
;   LENGTH: 2079 base pairs
;   TYPE: nucleic acid
;   STRANDEDNESS: single
;   TOPOLOGY: linear
; IMMEDIATE SOURCE:
;   LIBRARY: TILYNOT05
;   CLONE: 2809795
; SEQUENCE DESCRIPTION: SEQ ID NO: 2:
us-09-978-242-2

Query Match      100.0%; Score 2079; DB 9; Length 2079;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 2079; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 GGGCGCGCGCGCACCGGGAGCGCTCGGAGCGAGTGGAACTGGATCGGTTTCTG 60
Db 1 GGGCGCGCGCGCACCGGGAGCGCTCGGAGCGAGTGGAACTGGATCGGTTTCTG 60

Qy 61 CCAGGGCGGTGAGCTTCGGCCGGGCAATTTACACAGCTCCACTCGCGCGGACACAGGA 120
Db 61 CCAGGGCGGTGAGCTTCGGCCGGGCAATTTTACACAGCTCCACTCGCGCGGACACAGGA 120

Qy 121 GCAGCGACGCGCTTCCCGCAACCGCATACCATCGGACAGGATTTCTCGCCTCAGCC 180
Db 121 GCAGCGACGCGCTTCCCGCAACCGCATACCATCGGACAGGATTTCTCGCCTCAGCC 180

Qy 181 CAACGGGAGATCTCTGGAAACATGGCTACAGAAACATGTTAATGGAAATGGTACTGAAGA 240
Db 181 CAACGGGAGATCTCTGGAAACATGGCTACAGAAACATGTTAATGGAAATGGTACTGAAGA 240

Qy 241 GCCCATGATATCTTCTCGAGTTATCCATTACAGAAATTTTACAGATTCGTTGATGC 300
Db 241 GCCCATGATATCTTCTCGAGTTATCCATTACAGAAATTTTACAGATTCGTTGATGC 300

Qy 301 TGGTTTACACAGAAAGTTGCTGAAACACTAGATGAATTTACGTTGCGAGGCTAGTTGC 360
Db 301 TGGTTTACACAGAAAGTTGCTGAAACACTAGATGAATTTACGTTGCGAGGCTAGTTGC 360

Qy 361 ACATAGTGATTTAGATGAAGAGCTATTGAAGCTTTTAAAGAAATTTCAATGAAGACGGTGC 420
Db 361 ACATAGTGATTTAGATGAAGAGCTATTGAAGCTTTTAAAGAAATTTCAATGAAGACGGTGC 420

Qy 421 ATTGGCAGTCTTCAACAGTTTAAAGACAGTGATCTCTCATGTTTCAGAAACAAAGTGC 480
Db 421 ATTGGCAGTCTTCAACAGTTTAAAGACAGTGATCTCTCATGTTTCAGAAACAAAGTGC 480

Qy 481 CTTTTTATGTCGAGTCATGAAGCTTTACAGCAGAGAGAAACAAAGGGACCAAGTAGC 540
Db 481 CTTTTTATGTCGAGTCATGAAGCTTTACAGCAGAGAGAAACAAAGGGACCAAGTAGC 540

Qy 541 AGATTCTAGTAAAGACACAGATGAGGCAAAATTAAGGCACTCTTTGGAAAGAACAGGCTA 600
Db 541 AGATTCTAGTAAAGACACAGATGAGGCAAAATTAAGGCACTCTTTGGAAAGAACAGGCTA 600

Qy 601 CACACTTGATGTGACCATCTGGACAGAGAAAGTATGGAGGACCACTTCAGATTCGTTTGA 660
Db 601 CACACTTGATGTGACCATCTGGACAGAGAAAGTATGGAGGACCACTTCAGATTCGTTTGA 660

Qy 661 TTCAGTTCAGAGCCCTCTGTTGGCACTGAGATATTTCTGGGAAAGATCCCAAGAGATCT 720
Db 661 TTCAGTTCAGAGCCCTCTGTTGGCACTGAGATATTTCTGGGAAAGATCCCAAGAGATCT 720

Qy 721 ATTTGAGGATGAATCTGTTTCCATTTATTTGAGAAAGCTGACCTATATGGGATCTTCGTCT 780
Db 721 ATTTGAGGATGAATCTGTTTCCATTTATTTGAGAAAGCTGACCTATATGGGATCTTCGTCT 780

Qy 781 AATGATGATCCACTCACTGCTCAATAGAGGTTATCGGTTGTCATTTTGTACAAA 840
Db 781 AATGATGATCCACTCACTGCTCAATAGAGGTTATCGGTTGTCATTTTGTACAAA 840

841 AGAAGCAGCTCAGGAGCGCTGTTAAACTGTATATATCATGAAATTCGTTCTGGAAACA 900
841 AGAAGCAGCTCAGGAGCGCTGTTAAACTGTATATATCATGAAATTCGTTCTGGAAACA 900

901 TATTGTGCTCTGCATCTCAGTTGCGCAACAAATAGGCTTTTGTGGGCTCTATTCTTAAGAG 960
901 TATTGTGCTCTGCATCTCAGTTGCGCAACAAATAGGCTTTTGTGGGCTCTATTCTTAAGAG 960

961 TAAACCAAGGACAGATTTCTTGAAGAAATTTAGCAAAGTAAACAGAGGGCTTTACAGACGT 1020
961 TAAACCAAGGACAGATTTCTTGAAGAAATTTAGCAAAGTAAACAGAGGGCTTTACAGACGT 1020

1021 CATTTTATATACCAACCGGATGACAAAGAAAAAAGAAAAAAGAGGCTTTTGTCTTCAATA 1080
1021 CATTTTATATACCAACCGGATGACAAAGAAAAAAGAAAAAAGAGGCTTTTGTCTTCAATA 1080

1081 TGAAGATCAAAAACAGCTGCCAGCAAGGCTAGGTTAATGAGTGGTAAAGTCAAGGT 1140
1081 TGAAGATCAAAAACAGCTGCCAGCAAGGCTAGGTTAATGAGTGGTAAAGTCAAGGT 1140

1141 CTGGGGGAATGTTGGAACTGTTGAAATGGGCTGATCTCTATAGAAAGATCTCTGATCCTCAGGT 1200
1141 CTGGGGGAATGTTGGAACTGTTGAAATGGGCTGATCTCTATAGAAAGATCTCTGATCCTCAGGT 1200

1201 TATGGCAAAAGGTAAAGTGTCTGTTGTACGCAACCTTGGCAATATCTGTAAACAGAGAGAT 1260
1201 TATGGCAAAAGGTAAAGTGTCTGTTGTACGCAACCTTGGCAATATCTGTAAACAGAGAGAT 1260

1261 TTTAGAAAAGGCAATTTAGTCAGTTTGGGAAAATCTGGAAACAGAGTGAAGATTAAGATTA 1320
1261 TTTAGAAAAGGCAATTTAGTCAGTTTGGGAAAATCTGGAAACAGAGTGAAGATTAAGATTA 1320

1321 TCGGTTCAATTCATTTTGTAGTCAGCGAGATGCTGCTGCTCAAGGCTATCGAAGAAATGAATGG 1380
1321 TCGGTTCAATTCATTTTGTAGTCAGCGAGATGCTGCTGCTCAAGGCTATCGAAGAAATGAATGG 1380

1381 CAAAGACTTGGAGGAGAAAATATTGAAATTTGTTTTTGGCAACCCACCCAGATCAGAAAAG 1440
1381 CAAAGACTTGGAGGAGAAAATATTGAAATTTGTTTTTGGCAACCCACCCAGATCAGAAAAG 1440

1441 GAAAGAAAAGAAAGCTCAGAGGCAAGCAGCAAAAATCAATGTATGACGATTAAGTACTACTA 1500
1441 GAAAGAAAAGAAAGCTCAGAGGCAAGCAGCAAAAATCAATGTATGACGATTAAGTACTACTA 1500

1501 TTTATGTCACCTCATATGCCCTCCCAACAGAGGTCCAGGCGCTGGAGGTAGAGGTGG 1560
1501 TTTATGTCACCTCATATGCCCTCCCAACAGAGGTCCAGGCGCTGGAGGTAGAGGTGG 1560

1561 TTTATGATATCTCTCCAGATTTATTTATGATATGAAGATTTATTTATGTTTATGA 1620
1561 TTTATGATATCTCTCCAGATTTATTTATGATATGAAGATTTATTTATGTTTATGA 1620

1621 TTTACCAATACTATCTGTTGGATATGAAGATTCATATCTATGTTTATGAAGATTTTCAAGT 1680
1621 TTTACCAATACTATCTGTTGGATATGAAGATTCATATCTATGTTTATGAAGATTTTCAAGT 1680

1681 TCGAGCTAGAGGAGGCTGTTAGAGGACCAAGGGTGTCTCCATCCAGAGGTCTGTTG 1740
1681 TCGAGCTAGAGGAGGCTGTTAGAGGACCAAGGGTGTCTCCATCCAGAGGTCTGTTG 1740

1741 GGCTGCTCTCTCCCGGGTAGAGCCGGTTATTTCACAGAGAGGAGGCTCTGAGTCAAG 1800
1741 GGCTGCTCTCTCCCGGGTAGAGCCGGTTATTTCACAGAGAGGAGGCTCTGAGTCAAG 1800

1801 AGCGTTTCAGAGGAGGAGGCTCGAGGCGGCTCTGACCTGTTTACAAATGAAGACTGA 1860
1801 AGCGTTTCAGAGGAGGAGGCTCGAGGCGGCTCTGACCTGTTTACAAATGAAGACTGA 1860

1861 CTTGCTATGTTGGGATTTACACAGAACTTGCAGTGGAGTAACTGTAAGGAAATCAAGCAA 1920
1861 CTTGCTATGTTGGGATTTACACAGAACTTGCAGTGGAGTAACTGTAAGGAAATCAAGCAA 1920
```

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: January 29, 2004, 01:57:54 ; Search time 532 Seconds
(without alignments)

13976.470 Million cell updates/sec

Title: US-09-978-242-2

Perfect score: 2079

Sequence: 1 GGGCGCGCGCGCGCACCGG.....GTTATTAACTAAAGTACT 2079

Scoring table: IDENTITY NUC

Gapop 10.0 , Gapext 1.0

Searched: 2356869 seqs, 1788235258 residues

Total number of hits satisfying chosen parameters: 4713738

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

Published Applications NA:*

1: /cgn2_6/ptodata/1/pubpna/US07_PUBCOMB.seq.*
2: /cgn2_6/ptodata/1/pubpna/PCT_NEW_PUB.seq.*
3: /cgn2_6/ptodata/1/pubpna/US06_NEW_PUB.seq.*
4: /cgn2_6/ptodata/1/pubpna/US06_PUBCOMB.seq.*
5: /cgn2_6/ptodata/1/pubpna/US07_NEW_PUB.seq.*
6: /cgn2_6/ptodata/1/pubpna/PCTUS_PUBCOMB.seq.*
7: /cgn2_6/ptodata/1/pubpna/US08_NEW_PUB.seq.*
8: /cgn2_6/ptodata/1/pubpna/US08_PUBCOMB.seq.*
9: /cgn2_6/ptodata/1/pubpna/US09A_PUBCOMB.seq.*
10: /cgn2_6/ptodata/1/pubpna/US09B_PUBCOMB.seq.*
11: /cgn2_6/ptodata/1/pubpna/US09C_PUBCOMB.seq.*
12: /cgn2_6/ptodata/1/pubpna/US09_NEW_PUB.seq.*
13: /cgn2_6/ptodata/1/pubpna/US09A_PUBCOMB.seq.*
14: /cgn2_6/ptodata/1/pubpna/US10A_PUBCOMB.seq.*
15: /cgn2_6/ptodata/1/pubpna/US10B_PUBCOMB.seq.*
16: /cgn2_6/ptodata/1/pubpna/US10_NEW_PUB.seq.*
17: /cgn2_6/ptodata/1/pubpna/US60_NEW_PUB.seq.*
18: /cgn2_6/ptodata/1/pubpna/US60_PUBCOMB.seq.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	2079	100.0	2079	9	US-09-978-242-2
2	1844	88.7	4231	10	US-09-925-300-475
3	1826.2	87.8	3452	10	US-09-821-687-3
4	1813.4	87.2	2369	15	US-10-106-698-1042
5	882	42.4	2673	11	US-09-374-046A-127
6	882	42.4	3159	14	US-10-071-766-85
7	871.6	41.9	2671	11	US-09-822-846-199
8	839.8	40.4	2073	15	US-10-198-846-13811
9	833.6	40.1	2701	15	US-10-106-698-334
10	730	35.1	1024	9	US-09-925-299-68
11	730	35.1	1024	11	US-09-925-299-68
12	498.8	24.0	612	10	US-09-879-536-654
13	487	23.4	1907	9	US-09-822-849A-263
14	434.8	20.9	1176	15	US-10-198-846-9841
15	425	20.4	464	9	US-09-864-761-238

16	414.4	19.9	418	10	US-09-796-692-7702	Sequence 7702, Ap
17	414.4	19.9	418	12	US-10-057-475B-7702	Sequence 7702, Ap
18	414.4	19.9	418	12	US-10-154-884B-7702	Sequence 7702, Ap
19	414.4	19.9	418	15	US-10-040-862-7702	Sequence 7702, Ap
20	395.8	19.0	483	10	US-09-821-687-1	Sequence 1, Appli
21	391	18.8	403	11	US-09-918-998-8445	Sequence 8445, Ap
c	367	17.7	2715	15	US-10-106-698-1041	Sequence 1041, Ap
23	354.6	17.1	747	9	US-09-910-943-643	Sequence 643, App
24	287.8	13.8	591	15	US-10-106-698-2180	Sequence 2180, Ap
25	275.2	13.2	700	9	US-09-910-943-301	Sequence 301, App
26	272.8	13.1	382	9	US-09-864-761-17060	Sequence 17060, A
27	257.6	12.4	2022	13	US-09-949-029-99	Sequence 99, Appl
28	253.2	12.2	1992	14	US-10-098-841-310	Sequence 310, App
29	249.2	12.0	408	11	US-09-918-995-18621	Sequence 18621, A
30	248.4	11.9	500	11	US-09-918-995-24031	Sequence 24031, A
31	215	10.3	387	15	US-10-198-846-5910	Sequence 5910, Ap
32	212	10.2	368	10	US-09-960-352-9809	Sequence 9809, Ap
33	192.6	9.3	405	11	US-09-918-995-33148	Sequence 33148, A
34	180.4	8.7	429	10	US-09-738-973-414	Sequence 414, App
35	180.4	8.7	429	10	US-09-854-133-414	Sequence 414, App
36	180.4	8.7	429	15	US-10-144-649A-414	Sequence 414, App
37	165.2	7.9	480	15	US-10-198-846-7972	Sequence 7972, Ap
38	155.2	7.5	608	9	US-09-764-887-48	Sequence 48, Appl
39	155.2	7.5	608	15	US-10-073-961-48	Sequence 48, Appl
40	153.2	7.4	589	13	US-10-029-386-12112	Sequence 12112, A
41	141.6	6.8	286	10	US-09-867-701-12	Sequence 12, Appl
42	120.2	5.8	761	9	US-09-910-943-415	Sequence 415, App
43	110.2	5.3	220	13	US-10-029-386-25812	Sequence 25812, A
c	90.6	4.4	626	13	US-10-027-632-291227	Sequence 291227, A
45	90.6	4.4	626	13	US-10-027-632-291228	Sequence 291228, A

ALIGNMENTS

RESULT 1

US-09-978-242-2
; Sequence 2, Application US/09978242
; Patent No. US20020098566A1
; GENERAL INFORMATION:
; APPLICANT: Bandman, Olga
; Yue, Henry
; Corley, Neil C.
; Shah, Purvi
; TITLE OF INVENTION: HUMAN NUCLEOLIN-LIKE PROTEIN
; NUMBER OF SEQUENCES: 3
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Incyte Pharmaceuticals, Inc.
; STREET: 3174 Porter Drive
; CITY: Palo Alto
; STATE: CA
; COUNTRY: USA
; ZIP: 94304
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette
; COMPUTER: IBM Compatible
; OPERATING SYSTEM: DOS
; SOFTWARE: FastSeq for Windows Version 2.0
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/978,242
; FILING DATE: 15-Oct-2001
; CLASSIFICATION: <Unknown>
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 09/241,333
; FILING DATE: <Unknown>
; APPLICATION NUMBER: 08/990,114
; FILING DATE: <Unknown>
; ATTORNEY/AGENT INFORMATION:
; NAME: Billings, Lucy J.
; REGISTRATION NUMBER: 36,749
; REFERENCE/DOCKET NUMBER: PF-0451 US
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 650-855-0555

REFERENCE
AUTHORS

3 Shibata, K., Itoh, M., Aizawa, K., Nagao, S., Sasaki, N., Carninci, P.,
Konno, H., Akiyama, J., Nishi, K., Kitsuai, T., Tashiro, H., Itoh, M.,
Sumi, N., Ishii, Y., Nakamura, S., Hazama, M., Nishine, T., Harada, A.,
Yamamoto, R., Inoue, K., Togawa, Y., Izawa, M., Ohara, E., Watahiki, M.,
Fujiwara, S., Inoue, K., Ozawa, K., Tanaka, T., Matsuyama, S., Kawai, J.,
Okazaki, Y., Muramatsu, M., Inoue, Y., Kira, A., and Hayashizaki, Y.
RIKEN integrated sequence analysis (RISA) system--384-format
sequencing pipeline with 384 multicapillary sequencer
Genome Res. 10 (11), 1757-1771 (2000)

TITLE

JOURNAL
MEDLINE
PUBMED

20530913
11076861

REFERENCE
AUTHORS

4 Kawai, J., Shinagawa, A., Shibata, K., Yoshino, M., Itoh, M., Ishii, Y.,
Arakawa, T., Hara, A., Fukunishi, Y., Konno, H., Adachi, J., Fukuda, S.,
Aizawa, K., Izawa, M., Nishi, K., Kiyosawa, H., Kondo, S., Yamanaka, I.,
Saito, T., Okazaki, Y., Gojobori, T., Bono, H., Kasukawa, T., Saito, R.,
Kadota, K., Matsuda, H., Ashburner, M., Batalov, S., Casavant, T.,
Fleischmann, W., Gaasterland, T., Gissi, C., King, B., Kochiwa, H.,
Kuehl, P., Lewis, S., Matsuo, Y., Nikaide, I., Pesole, G.,
Quackenbush, J., Schriml, L.M., Stauber, F., Suzuki, R., Tomita, M.,
Wagner, L., Washio, T., Sakai, K., Okido, T., Furuno, M., Aono, H.,
Baldarelli, R., Barsh, G., Blake, J., Boffelli, D., Bojunga, N.,
Carninci, P., de Bonaldo, M.F., Brownstein, M.J., Bult, C.,
Fletcher, C., Fujita, M., Gariboldi, M., Gustincich, S., Hill, D.,
Hofmann, M., Hume, D.A., Kaniwa, M., Lee, N.H., Lyons, P.,
Marchionni, L., Mashima, J., Mazzarelli, J., Mombauts, P., Nordone, P.,
Ring, B., Ringwald, M., Rodriguez, I., Sakamoto, N., Sasaki, H.,
Sato, K., Schonbach, C., Seya, T., Shibata, Y., Storch, K.F., Suzuki, H.,
Toyo-oka, K., Wang, K.H., Weitz, C., Whittaker, C., Wilming, L.,
Wynshaw-Boris, A., Yoshida, K., Hasegawa, Y., Kawaji, H., Kohtsuki, S.,
and Hayashizaki, Y.
Functional annotation of a full-length mouse cDNA collection
Nature 409 (6821), 685-690 (2001)

TITLE

JOURNAL
MEDLINE
PUBMED

21085660
11217851

REFERENCE
AUTHORS

5 The FANTOM Consortium and the RIKEN Genome Exploration Research
Group Phase I & II Team.
Analysis of the mouse transcriptome based on functional annotation
of 60,770 full-length cDNAs
Nature 420, 563-573 (2002)

TITLE

JOURNAL
MEDLINE
PUBMED

6 (bases 1 to 3838)

REFERENCE
AUTHORS

Adachi, J., Aizawa, K., Akimura, T., Arakawa, T., Bono, H., Carninci, P.,
Fukuda, S., Furuno, M., Hangaki, T., Hara, A., Hashizume, M.,
Hayashida, K., Hayatsu, N., Hiramoto, K., Hiraoka, T., Hirozane, T.,
Hori, F., Imotani, K., Ishii, Y., Itoh, M., Kagawa, I., Kasukawa, T.,
Kato, H., Kawai, J., Kojima, Y., Kondo, S., Konno, H., Kouda, M.,
Koya, S., Kurihara, C., Matsuyama, T., Miyazaki, A., Murata, M.,
Nakamura, M., Nishi, K., Nomura, K., Numazaki, R., Ohno, M., Ohsato, N.,
Okazaki, Y., Saito, R., Saitoh, H., Sakai, C., Sakai, K., Sakazume, N.,
Sano, H., Sasaki, D., Shibata, K., Shinagawa, A., Shiraki, T.,
Sogabe, Y., Tagami, M., Tagawa, A., Takahashi, F., Takaku-Akahira, S.,
Takeda, Y., Tanaka, T., Tomaru, A., Toya, T., Yasunishi, A.,
Muramatsu, M., and Hayashizaki, Y.
Direct Submission

TITLE

JOURNAL

Submitted (16-JUL-2001) Yoshihide Hayashizaki, The Institute of
Physical and Chemical Research (RIKEN), Laboratory for Genome
Exploration Research Group, RIKEN Genomic Sciences Center (GSC),
RIKEN Yokohama Institute, 1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama,
Kanagawa 230-0045, Japan [E-mail: genome-res@gsc.riken.go.jp,
URL: http://genome.gsc.riken.go.jp/, Tel: 81-45-503-9222,
Fax: 81-45-503-9216]

COMMENT

cDNA library was prepared and sequenced in Mouse Genome
Encyclopedia Project of Genome Exploration Research Group in Riken
Genomic Sciences Center and Genome Science Laboratory in RIKEN.
Division of Experimental Animal Research in Riken contributed to
prepare mouse tissues.
Please visit our web site for further details.
URL: http://genome.gsc.riken.go.jp/
URL: http://fantom.gsc.riken.go.jp/
Location/Qualifiers

FEATURES

source	1..3838	/organism="Mus musculus"
		/mol_type="mRNA"
		/strain="C57BL/6J"
		/db_xref="PANTOM:DB:9430046J23"
		/clone="9430046J23"
		/tissue_type="embryonic body between diaphragm region and neck"
		/clone_lib="RIKEN full-length enriched mouse cDNA library"
		/dev_stage="12 days embryo"
CDS	167..1855	/note="unnamed protein product; NSI-associated protein 1 (MGDI [MGJ:1891690, GB AB035725, evidence: BLASTN, 99%, match=3107])"
		/protein_id="BAC28852.1"
		/db_xref="GI:26330244"
		/translation="WATEHNGNGTEPEMDTTSVAIHSENFOTLLDAGLPQKVAEKLD EIVAGLVASHDLDERAIEALKEFNEDEGALVLOQFKDSLSHVQNKSAFLCQVMTY ROREKGTQVADSSSGPDEAKIKALLERTGYTLDTVTGQKYGPPDSVYSQOQSV GTEIFVKIIPROLFEDFVLPFEKAGPIWDLRLMDPLTGLARGVAFVFTCTKEAAGE AVKLYNNHBSIRSGKIGVCISVANNKLFVGSIPKSKTEQILIEFSKYTEGLTDVLY AKVKVLFVRNLANTVTEILEKFSQFGKLERVKLKYAFIHFDRDGAIVKAMEBN KDLEGENTIEIVFAKPPDKRKAQROAKNQMYDDYYFYYPHPMPPTGRGRGG RGGYGPDPYGYEDYDYDYDHYHNGYEDPYGYEDFQVARGRGRGRGARGAAP SRGGAAPPRGRAGYSQSGPSARGVARGAQAQOQGRGQGGKGVGEAGPDLQ"
BASE COUNT	1164 a 662 c 843 g 1169 t	
ORIGIN		
	Query Match	89.0%; Score 1850.4; DB 11; Length 3838;
	Best Local Similarity	94.4%; Pred. No. 0;
	Matches 1966; Conservative	0; Mismatches 76; Indels 40; Gaps 3;
Qy	38	GTGGAATTCGGATCGGGTTTCTGCCAGCGCGGTGAGCTTCGGCGCGCATTTTCAACAGC 97
Db	2	GTGGAATTCGGATCGGGTTTCTGCCAGCGCGGTGAGCTTCGGCGCGCATTTTCAACAGC 61
Qy	98	TCACATCGCGCGGACACAGGAGGAGCGGTTTCCCGCAACCGGATACCATCG 157
Db	62	TCACATCGCGCGGACACAGGAGGAGCGGTTTCCCGCAACCGGATACCATCG 121
Qy	158	GACAGATTCTCGGCTCAGCCCAACGGGAGATCTCTGGAACATGGCTACAGAACAT 217
Db	122	GACAGATTCTCGGCTCAGCCCAACGGGAGATCTCTGGAACATGGCTACAGAACAT 181
Qy	218	GTTAATGGAAATGGTACTGAAGAGCCCATGGATACCTACTCTCGAGTTATCCATTGAGAA 277
Db	182	GTTAATGGAAATGGTACTGAAGAGCCCATGGATACCTACTCTCGAGTTATCCATTGAGAA 241
Qy	278	AATTTTCAGACATTCGTTGATGCTGGTTTACACAGAAAGTTCTGTAAGAACTAGATGAA 337
Db	242	AATTTTCAGACATTCGTTGATGCTGGTTTACACAGAAAGTTCTGTAAGAACTAGATGAA 301
Qy	338	ATTTCAGTTCAGCGGCTAGTTGCACATAGTGTATTTAGTGAAGAGCTATTGAAGCTTTA 397
Db	302	ATTTCAGTTCAGCGGCTAGTTGCACATAGTGTATTTAGTGAAGAGCTATTGAAGCTTTA 361
Qy	398	AAAGAAATTCAGAAAGCGGTGATTCGGAGTTTCTTCAACAGTTTAAAGACAGTGATCTC 457
Db	362	AAAGAGTTCAATGAAGAGCGGATTCGGAGTTTCTTCAACAGTTTAAAGACAGTGATCTC 421
Qy	458	TCTCATGTTTCAGAAACAAAGTGCCTTTTATGTGGAGTCAATGAAGCTTACAGCGCAGA 517
Db	422	TCTCATGTTTCAGAAACAAAGTGCCTTTTATGTGGAGTCAATGAAGCTTACAGCGCAGA 481
Qy	518	GAAGAAACAGGAGACCAAGTAGCAGATTCTAGTAAAGGACAGATGAGGCAAAATTAAG 577
Db	482	GAAGAAACAGGAGACCAAGTAGCAGATTCTAGTAAAGGACAGATGAGGCAAAATTAAG 541
Qy	578	GCATCTTGGAAAGAACAGGCTACACATTTGATGTGACCACTGGACAGAGGAAGTATGGA 637

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: January 29, 2004, 00:06:34 ; Search time 2986 Seconds
(without alignments)
16921.971 Million cell updates/sec

Title: US-09-978-242-2

Perfect score: 2079

Sequence: 1 GGGCGCGCGCGCACCGGG.....GTTATTAACTAAAGCTACT 2079

Scoring table: IDENTITY NUC

Gapop 10.0 , Gapext 1.0

Searched: 22781392 seqs, 12152238056 residues

Total number of hits satisfying chosen parameters: 45562784

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

EST:*

1: em_estba:*

2: em_esthum:*

3: em_estin:*

4: em_estmu:*

5: em_estov:*

6: em_estpl:*

7: em_estro:*

8: em_htc:*

9: gb_estl:*

10: gb_est2:*

11: gb_est3:*

12: gb_est4:*

13: gb_est5:*

14: gb_est6:*

15: em_estfun:*

16: em_estom:*

17: em_gss_hum:*

18: em_gss_inv:*

19: em_gss_pln:*

20: em_gss_vrt:*

21: em_gss_fun:*

22: em_gss_mam:*

23: em_gss_mus:*

24: em_gss_pro:*

25: em_gss_rod:*

26: em_gss_phg:*

27: em_gss_vrl:*

28: gb_gss1:*

29: gb_gss2:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	1850.4	89.0	3838	11 AK034845	AK034845 Mus muscu
2	1792.8	86.2	2141	11 AK076026	AK076026 Mus muscu
3	1787.4	86.0	2030	11 AK077588	AK077588 Mus muscu
4	1537	73.9	4064	11 AK083398	AK083398 Mus muscu

5	1494.8	71.9	3289	11 AK078158	AK078158 Mus muscu
6	1386	66.7	1405	11 BC021932	BC021932 Homo sapi
7	1353.6	65.1	1361	11 BC024283	BC024283 Homo sapi
8	1343.6	64.6	2618	11 BC019360	BC019360 Homo sapi
9	1331	64.0	1359	11 BC009176	BC009176 Homo sapi
10	1327.4	63.8	1343	11 BC015575	BC015575 Homo sapi
11	983.8	47.3	1201	9 AL541211	AL541211 AL541211
12	975.2	46.9	1201	9 AL541923	AL541923 AL541923
13	921.6	44.3	1201	9 AL514248	AL514248 AL514248
14	884.2	42.5	2663	11 BC026850	BC026850 Mus muscu
15	883	42.5	1144	12 BM466029	BM466029 AGENCOURT
16	866.6	41.7	946	13 BQ686098	BQ686098 AGENCOURT
17	862.8	41.5	1080	12 BM809311	BM809311 AGENCOURT
18	844.6	40.6	905	13 BQ945048	BQ945048 AGENCOURT
19	838.2	40.3	1201	13 BX420656	BX420656 BX420656
20	836.6	40.2	909	13 BU506813	BU506813 AGENCOURT
21	831.8	40.0	868	13 BU194176	BU194176 AGENCOURT
22	829.2	39.9	910	13 BU183896	BU183896 AGENCOURT
23	826.8	39.8	879	12 BM451217	BM451217 AGENCOURT
24	807.6	38.8	929	13 BU845727	BU845727 AGENCOURT
25	768	36.9	1063	12 BM543159	BM543159 AGENCOURT
26	757.8	36.5	856	13 BQ680184	BQ680184 AGENCOURT
27	756.6	36.4	921	13 BU505018	BU505018 AGENCOURT
28	744.6	35.8	878	13 BQ650432	BQ650432 AGENCOURT
29	740.8	35.6	903	13 BQ680994	BQ680994 AGENCOURT
30	740.2	35.6	933	13 BQ950249	BQ950249 AGENCOURT
31	736.2	35.4	920	13 BQ959242	BQ959242 AGENCOURT
32	729	35.1	988	9 AV298691	AV298691 AV298691
33	728.6	35.0	891	14 CB182506	CB182506 AGENCOURT
34	721.2	34.7	826	10 BE795387	BE795387 601592924
35	719	34.6	777	13 BQ443866	BQ443866 UI-M-EW0-
36	715.6	34.4	1032	10 BE407142	BE407142 601301856
37	712.8	34.3	950	10 BG026689	BG026689 602233533
38	710.6	34.2	798	12 BI558562	BI558562 603240561
39	709.8	34.1	823	10 BG282140	BG282140 602403281
40	704.6	33.9	772	13 BQ571781	BQ571781 UI-M-FC0-
41	700.4	33.7	1088	12 BI661316	BI661316 603304285
42	700.2	33.7	860	10 BE560849	BE560849 601346126
43	698.8	33.6	895	12 BI558142	BI558142 603240061
44	697.8	33.6	856	13 BU191822	BU191822 AGENCOURT

ALIGNMENTS

RESULT 1	AK034845	AK034845	3838 bp	linear	HTC 05-DEC-2002
LOCUS	Mus musculus	Mus musculus	12 days embryo	embryonic body	between diaphragm region
DEFINITION	and neck CDNA, RIKEN full-length enriched library, clone:9430046J23	Product: NSI-associated protein 1, full insert sequence.			
ACCESSION	AK034845	AK034845.1	GI:26330243		
VERSION	AK034845	HTC; CAP trapper.			
KEYWORDS	HTC; CAP trapper.				
SOURCE	Mus musculus (house mouse)				
ORGANISM	Mus musculus				
REFERENCE	1				
AUTHORS	Carninci, P., Shibata, Y., Hayatsu, N., Sugahara, Y., Shibata, K.,				
TITLE	High-efficiency full-length cDNA cloning				
JOURNAL	Meth. Enzymol. 303, 19-44 (1999)				
MEDLINE	99279253				
PUBMED	10349636				
REFERENCE	2				
AUTHORS	Carninci, P., Shibata, Y., Hayatsu, N., Sugahara, Y., Shibata, K.,				
TITLE	Normalization and subtraction of cap-trapper-selected cDNAs to				
JOURNAL	Genome Res. 10 (10), 1617-1630 (2000)				
MEDLINE	20499374				
PUBMED	11042159				